



Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>
Handle ID: <http://hdl.handle.net/10985/18921>

To cite this version :

Alain LERAT - An efficient high-order compact scheme for the unsteady compressible Euler and Navier–Stokes equations - Journal of Computational Physics - Vol. 322, p.365-386 - 2016

Any correspondence concerning this service should be sent to the repository

Administrator : scienceouverte@ensam.eu



An efficient high-order compact scheme for the unsteady compressible Euler and Navier–Stokes equations

A. Lerat

DynFluid Lab., Arts et Metiers ParisTech, 151 Boulevard de l'Hopital, 75013 Paris, France

A B S T R A C T

Residual-Based Compact (*RBC*) schemes approximate the 3-D compressible Euler equations with a 5th- or 7th-order accuracy on a $5 \times 5 \times 5$ -point stencil and capture shocks pretty well without correction. For unsteady flows however, they require a costly algebra to extract the time-derivative occurring at several places in the scheme. A new high-order time formulation has been recently proposed [13] for simplifying the *RBC* schemes and increasing their temporal accuracy. The present paper goes much further in this direction and deeply reconsiders the method. An avatar of the *RBC* schemes is presented that greatly reduces the computing time and the memory requirements while keeping the same type of successful numerical dissipation. Two and three-dimensional linear stability are analyzed and the method is extended to the 3-D compressible Navier–Stokes equations. The new compact scheme is validated for several unsteady problems in two and three dimension. In particular, an accurate DNS at moderate cost is presented for the evolution of the Taylor–Green Vortex at Reynolds 1600 and Prandtl 0.71. The effects of the mesh size and of the accuracy order in the approximation of Euler and viscous terms are discussed.

Keywords:

Compact schemes

High order

Unsteady compressible flows

Taylor–Green Vortex

1. Introduction

Among high-order methods for computing compressible flows on structured meshes, compact schemes are attractive because of their narrow grid-stencil that significantly reduces the truncation error for a given accuracy-order and makes the treatment of deforming meshes and boundary conditions easier. Compact schemes for compressible flows have been mainly developed as centered approximations in space, notably in the works by Lele [1], Cockburn and Shu [2], Yee [3] and Visbal and Gaitonde [4], with numerical dissipation based on artificial viscosities, numerical filters or limiters. Upwind compact schemes have also been proposed by Tolstykh [5] and Fu and Ma [6]. Another interesting option is the use of Residual-Based Compact (*RBC*) schemes. In a such a scheme, the consistent part and the numerical dissipation are expressed only in terms of compact approximations of the complete residual, *i.e.* the sum of the terms in the governing equations including the time derivative. These compact approximations are deduced from Padé formulas in which the inverse operators are eliminated. On a Cartesian mesh, a *RBC* scheme can approximate a hyperbolic system of conservation laws in d -dimension with a 5th- or 7th-order accuracy on a 5^d -point stencil and capture shocks pretty well without correction. Description and analysis of these schemes can be found in [7–13]. A related approach developed on unstructured meshes is the residual-distribution method of Abgrall, Deconinck and Ricchiuto [14–18] in which the residuals are distributed to the nodes of triangles or tetrahedrons.

E-mail address: alain.lerat@ensam.eu

A peculiarity of the *RBC* schemes is the multiple occurrence of the time derivative in the scheme ($\partial w/\partial t$ occurs $d + 1$ times in d -dimension). Besides, due to compactness, discrete spatial-operators are applied to each time-derivative. In the first applications of the *RBC* schemes to unsteady problems (see [19,10,11,20]), the time formulation was based on the Gear three-level method which is *A*-stable. This approach is efficient for computing compressible flows in steady and slow unsteady regimes. However, it requires an iterative method to advance the solution (dual-time approach or Newton sub-iterations) and its time accuracy is limited to order 2, which is not sufficient for some unsteady applications. This is the reason why an explicit time-formulation of high accuracy has been proposed in [13]. Since a direct extraction of the time-derivative requires the solution of a large linear algebraic-system, an approximate space-factorization of the time operator has been done with a correction to preserve rigorously the high order in space. This leads to the solution of simple linear systems that are block tridiagonal for *RBC* schemes spatially-accurate at order 3 or 5 and block pentadiagonal for the *RBC* scheme of order 7. Then, an *ABM* time-integration of order q (explicit combination of Adams–Bashforth and Adams–Moulton methods) is used to advance the solution. The *ABM* method offers the advantage of only requiring two computations per time-step of the flux balance and of the time-derivative extraction, whatever the order q .

In the present paper we go much further in the quest of simplifying the algorithm and reducing the CPU-time for the same accuracy level. We construct an avatar of the *RBC* schemes by dissociating the computation of the numerical dissipation from that of the consistent part. Doing so, the residual-based feature is partially lost, but the numerical dissipation of the new scheme discretizes the same partial derivative operator as the *RBC* scheme and only requires the solution of ordinary tridiagonal (up to order 5) or pentadiagonal linear systems without the need for a space factorization. The time-integration used for this simple scheme is the optimized low-storage Runge Kutta method *RKo6* of Bogey and Bailly [21].

The paper is organized as follows. Section 2 presents the various Padé compact approximations used in this paper and recalls the construction of the *RBC* scheme and the specific features of its numerical dissipation for solving a hyperbolic system of conservation laws in two-dimension. It also recalls the direct extraction of the time derivative with the factorization technique and the *RBC* – *ABM* schemes. Section 3 describes the new compact scheme for the Euler equations and its combination with *RKo6* in which the numerical dissipation is taken into account at the last stage only. The new form of numerical dissipation is compared to that of non-compact upwind schemes of high order. A multidimensional analysis of linear stability of the overall algorithm is presented for the various spatial orders and proves stability for *CFL* numbers equal to or slightly greater than one. Section 4 presents the extension of the new compact scheme of order 5 to the compressible Navier–Stokes equations. To discretize the second derivatives with variable coefficients on a 5-point stencil, we do not apply twice the Padé approximation of a first derivative, but use specific Padé formulas. The scheme here described is spatially accurate at order 5 for the Euler fluxes and at order 4 for general viscous fluxes (*E5V4* scheme) and only requires the solution of tridiagonal linear systems with constant coefficients. In Section 5, the new scheme is applied to several test-problems. We first consider the advection of a vortex during a long time in the horizontal or diagonal direction. These 2-D Euler problems allow a validation of the accuracy order of *E5* – *RKo6* for several *CFL* numbers and a measure of the important reduction in computing time over the *RBC5* – *ABM4* scheme. Then the 3-D scalar problem of the diagonal advection of a spherical Gaussian-shape is computed by the new scheme in order to confirm the accuracy order in 3D as well as the stability domain. Successful computations are done up to the *CFL* theoretical-limit of $2/\sqrt{3}$. The next problem is a direct numerical simulation of the Taylor–Green Vortex at $Re = 1600$ and $Pr = 0.71$. Several accuracy orders are considered for the Euler and viscous terms on a series of meshes. Enstrophy evolutions are compared to a reference solution and the CPU-costs are presented. The *E5V4* – *RKo6* scheme produces a very accurate solution on a 256^3 mesh for a moderate CPU-time. Finally, an Euler computation of shock–vortex interaction allows a first assessment of the shock-capturing capabilities of the new compact scheme, still used without any kind of correction. Conclusions are drawn and future work is planned in Section 6.

2. RBC schemes

2.1. Reminder on compact approximations

A derivative $\partial f/\partial x$ can be approximated at any order on a regular mesh $x_j = j \delta x$ using two simple discrete operators, i.e. a difference and an average over one mesh interval:

$$(\delta f)_{j+\frac{1}{2}} = f_{j+1} - f_j \quad (\mu f)_{j+\frac{1}{2}} = \frac{1}{2}(f_{j+1} + f_j)$$

where j is an integer or half integer. For instance:

$$(\delta \mu f)_j = (\mu f)_{j+\frac{1}{2}} - (\mu f)_{j-\frac{1}{2}} = \frac{1}{2}(f_{j+1} - f_{j-1})$$

$$(\delta^2 f)_j = (\delta \delta f)_j = f_{j+1} - 2f_j + f_{j-1}$$

$$(\delta^4 f)_j = f_{j+2} - 4f_{j+1} + 6f_j - 4f_{j-1} + f_{j-2}$$

Using δ and μ , we can easily write the 9-point centered formula for the 8th-order approximation of $\partial f/\partial x$:

Table 1Error-coefficient K_{2p} of the approximation of a derivative at order $2p = 4, 6, 8$.

Error coefficient	K_4	K_6	K_8
Non-compact	$-\frac{1}{30}$	$\frac{1}{140}$	$-\frac{1}{630}$
Compact	$-\frac{1}{180}$	$\frac{1}{2100}$	$-\frac{1}{44100}$

$$\left(\frac{\partial f}{\partial x}\right)_j = \left[\left(I - \frac{1}{6}\delta^2 + \frac{1}{30}\delta^4 - \frac{1}{140}\delta^6 \right) \frac{\delta \mu f}{\delta x} \right]_j + \mathcal{O}(\delta x^8) \quad (1)$$

where I is the identity operator. This writing also contains the approximations of lower orders. Indeed, by successively removing the terms $-\frac{1}{140}\delta^6$, $\frac{1}{30}\delta^4$ and $-\frac{1}{6}\delta^2$, we respectively get the approximations of order 6, 4 and 2 on a 7-, 5- and 3-point stencil.

To obtain the same accuracy orders on a shorter stencil, we can use a Padé compact approximation of the derivative, briefly called a Padé derivative, that is:

$$\left(\frac{\partial f}{\partial x}\right)_j = \left(\frac{\nabla f}{\delta x}\right)_j + \mathcal{O}(\delta x^{2p}) \quad (2)$$

with

$$\nabla = D^{-1}(I + a\delta^2)\delta\mu, \quad D = I + b\delta^2 + c\delta^4$$

where a , b and c are scalar parameters. The above formula involves 5 points at most. However, to extract the Padé derivative $\nabla f/\delta x$, we have to solve the linear algebraic system:

$$\left(D \frac{\nabla f}{\delta x}\right)_j = \left[(I + a\delta^2) \frac{\delta \mu f}{\delta x} \right]_j \quad (3)$$

The matrix associated to the D -operator is pentadiagonal in general.

The Padé derivative is second-order accurate at least. It becomes of fourth order at least if $a = b - \frac{1}{6}$. By choosing

$$a = 0, \quad b = \frac{1}{6}, \quad c = 0 \quad (4)$$

both sides of (3) have a 3-point stencil.

The Padé derivative is accurate at order 6 at least if the parameters satisfy

$$a = \frac{1}{30} + 6c, \quad b = \frac{1}{5} + 6c \quad (5)$$

By choosing $c = 0$, the D -operator has a 3-point stencil, which leads to a tridiagonal linear system.

The Padé derivative is accurate at order 8 for

$$a = \frac{5}{42}, \quad b = \frac{2}{7}, \quad c = \frac{1}{70} \quad (6)$$

The short grid-stencil of compact formulas makes the treatment of deforming meshes and boundary conditions easier. In addition, it significantly reduces the error. Indeed, the truncation error of the Padé derivative of order $2p = 4, 6, 8$ is:

$$\left(\frac{\nabla f}{\delta x} - \frac{\partial f}{\partial x}\right)_j = K_{2p} \delta x^{2p} \left(\frac{\partial^{2p+1} f}{\partial x^{2p+1}}\right)_j + \mathcal{O}(\delta x^{2p+2})$$

The truncation error of the non-compact formula (1) has the same form, but the error-coefficients K_{2p} of the compact formula are much smaller than those of the non-compact one, especially as the order increases (see Table 1). This is a strong argument in favor of compactness. Note that a similar result holds for second derivatives (see [9]).

To describe the numerical dissipation of the RBC schemes and to approximate the viscous fluxes, we also need Padé approximations at midpoints for a derivative and an averaging. They are defined as:

$$\left(\frac{\partial f}{\partial x}\right)_{j+\frac{1}{2}} = \left(\frac{\overline{\nabla} f}{\delta x}\right)_{j+\frac{1}{2}} + \mathcal{O}(\delta x^{2p}), \quad f_{j+\frac{1}{2}} = (\overline{\mu} f)_{j+\frac{1}{2}} + \mathcal{O}(\delta x^{2p}) \quad (7)$$

with

$$\overline{\nabla} = (D^\delta)^{-1}(I + a^\delta \delta^2) \delta, \quad \overline{\mu} = (D^\mu)^{-1}(I + a^\mu \delta^2) \mu$$

where

$$D^\delta = I + b^\delta \delta^2 + c^\delta \delta^4, \quad D^\mu = I + b^\mu \delta^2 + c^\mu \delta^4$$

The midpoint Padé approximations $\bar{\nabla}_{\delta x}$ and $\bar{\mu}$ are second-order accurate at least. They become of fourth order at least if

$$b^\delta - a^\delta = \frac{1}{24}, \quad b^\mu - a^\mu = \frac{1}{8} \quad (8)$$

They are accurate at order 6 at least if in addition

$$\frac{1}{24}b^\delta - c^\delta = \frac{3}{640}, \quad \frac{1}{8}b^\mu - c^\mu = \frac{3}{128} \quad (9)$$

To construct the numerical dissipation of the RBC schemes, midpoint approximations of orders 2, 4 and 6 will be sufficient. Besides we will choose $D^\mu = D^\delta$, i.e. $b^\delta = b^\mu$ and $c^\delta = c^\mu$, to obtain a great simplification in the dissipation term (vanishing of D^μ and D^δ). The simplest relevant sets of parameters are then the following ones. For 2nd-order:

$$a^\delta = a^\mu = b^\delta = b^\mu = c^\delta = c^\mu = 0 \quad (10)$$

For 4th-order, we satisfy (8) with $D^\mu = D^\delta$ and for simplicity $a^\mu = c^\mu = 0$, which gives

$$a^\delta = \frac{1}{12}, \quad a^\mu = 0, \quad b^\delta = b^\mu = \frac{1}{8}, \quad c^\delta = c^\mu = 0 \quad (11)$$

For 6th-order, we satisfy (8) and (9) with $D^\mu = D^\delta$, which gives

$$a^\delta = \frac{11}{60}, \quad a^\mu = \frac{1}{10}, \quad b^\delta = b^\mu = \frac{9}{40}, \quad c^\delta = c^\mu = \frac{3}{640} \quad (12)$$

Now concerning the approximation of viscous terms, the choice criterion is no longer $D^\mu = D^\delta$ but $c^\delta = c^\mu = 0$ to produce tridiagonal linear systems only. For 4th-order, we satisfy (8) with $c^\delta = c^\mu = 0$ and for simplicity $a^\mu = a^\delta = 0$, which gives:

$$a^\delta = a^\mu = 0, \quad b^\delta = \frac{1}{24}, \quad b^\mu = \frac{1}{8}, \quad c^\delta = c^\mu = 0 \quad (13)$$

For 6th-order, we satisfy (8) and (9) with $c^\delta = c^\mu = 0$, which gives

$$a^\delta = \frac{17}{240}, \quad a^\mu = \frac{1}{16}, \quad b^\delta = \frac{9}{80}, \quad b^\mu = \frac{3}{16}, \quad c^\delta = c^\mu = 0 \quad (14)$$

2.2. RBC space approximation on a 5×5 -point stencil

Consider the hyperbolic system of conservation laws:

$$\frac{\partial w}{\partial t} + \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} = 0 \quad (15)$$

where t is the time, x_1 and x_2 are the space coordinates, w is the state vector and $f_1 = f_1(w)$, $f_2 = f_2(w)$ are flux components depending smoothly on w . The Jacobian matrices of the flux are denoted $A_1 = df_1/dw$ and $A_2 = df_2/dw$.

System (15) is approximated in space on a uniform Cartesian mesh $(j_1 \delta x_1, j_2 \delta x_2)$ using a *residual-based-compact* (RBC) scheme. Such a scheme is a compact discrete form of

$$r = \frac{\delta x_1}{2} \frac{\partial(\Phi_1 r)}{\partial x_1} + \frac{\delta x_2}{2} \frac{\partial(\Phi_2 r)}{\partial x_2} \quad (16)$$

where r is the exact residual:

$$r = \frac{\partial w}{\partial t} + \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2}$$

and the right-hand side is a residual-based numerical dissipation in which Φ_1 and Φ_2 are matrices, each depending on the eigensystems of the Jacobian matrices A_1 and A_2 and on the ratio of the space steps. These matrices are defined in the sequel. They contain no tuning parameters or limiters.

To describe the space approximation of (16), we introduce the basic discrete operators in each space direction:

$$\begin{aligned} (\delta_1 v)_{j_1+\frac{1}{2}, j_2} &= v_{j_1+1, j_2} - v_{j_1, j_2} & (\delta_2 v)_{j_1, j_2+\frac{1}{2}} &= v_{j_1, j_2+1} - v_{j_1, j_2} \\ (\mu_1 v)_{j_1+\frac{1}{2}, j_2} &= \frac{1}{2}(v_{j_1+1, j_2} + v_{j_1, j_2}) & (\mu_2 v)_{j_1, j_2+\frac{1}{2}} &= \frac{1}{2}(v_{j_1, j_2+1} + v_{j_1, j_2}) \end{aligned}$$

where j_1 and j_2 are integers or half integers. These four discrete operators commute two by two.

Replacing the exact residual in Eq. (16) by different compact centered approximations \tilde{r}_{j_1, j_2} , $(\tilde{r}_1)_{j_1+\frac{1}{2}, j_2}$ and $(\tilde{r}_2)_{j_1, j_2+\frac{1}{2}}$, we obtain the residual-based compact scheme:

$$\tilde{r}_{j_1, j_2} = \frac{1}{2}[\delta_1(\Phi_1 \tilde{r}_1) + \delta_2(\Phi_2 \tilde{r}_2)]_{j_1, j_2} \quad (17)$$

In spite of appearances, the numerical dissipation in (17) is not simply of order one because (\tilde{r}_1) and (\tilde{r}_2) approximate the exact residual $r = 0$. Clearly, if \tilde{r}_{j_1, j_2} approximates r at order $2p$ and $(\tilde{r}_1)_{j_1+\frac{1}{2}, j_2}$, $(\tilde{r}_2)_{j_1, j_2+\frac{1}{2}}$ approximate r at order $2p - 2$, then the scheme (17) is accurate at order $2p - 1$ in space and denoted as RBC_{2p-1} .

We now restrict the scheme stencil to 5×5 points. For constructing the main residual \tilde{r}_{j_1, j_2} , we first replace the space derivatives in the exact residual r by Padé approximations $\frac{\nabla_1}{\delta x_1}$ and $\frac{\nabla_2}{\delta x_2}$ of order $2p$, where similarly as in 2.1:

$$\nabla_l = (D_l)^{-1}(I + a\delta_l^2) \delta_l \mu_l, \quad D_l = I + b\delta_l^2 + c\delta_l^4, \quad l = 1, 2 \quad (18)$$

Then we apply the operator $D_1 D_2$ to all the terms in the residual and obtain

$$\tilde{r}_{j_1, j_2} = \left[D_1 D_2 \frac{\partial w}{\partial t} + D_2(I + a\delta_1^2) \frac{\delta_1 \mu_1 f_1}{\delta x_1} + D_1(I + a\delta_2^2) \frac{\delta_2 \mu_2 f_2}{\delta x_2} \right]_{j_1, j_2} \quad (19)$$

For constructing the two residuals in numerical dissipation $(\tilde{r}_1)_{j_1+\frac{1}{2}, j_2}$ and $(\tilde{r}_2)_{j_1, j_2+\frac{1}{2}}$, we use Padé approximations at midpoints. More precisely, for $(\tilde{r}_1)_{j_1+\frac{1}{2}, j_2}$ we first replace $\frac{\partial f_1}{\partial x_1}$ by the Padé derivative at midpoints $\frac{\bar{\nabla}_1}{\delta x_1}$ and we apply the Padé average at midpoints $\bar{\mu}_1$ to $\frac{\partial w}{\partial t}$ and $\frac{\partial f_2}{\partial x_2}$, which gives

$$\left[\bar{\mu}_1 \frac{\partial w}{\partial t} + \frac{\bar{\nabla}_1 f_1}{\delta x_1} + \bar{\mu}_1 \frac{\partial f_2}{\partial x_2} \right]_{j_1+\frac{1}{2}, j_2}$$

with

$$\bar{\nabla}_l = (D_l^\delta)^{-1}(I + a^\delta \delta_l^2) \delta_l, \quad \bar{\mu}_l = (D_l^\mu)^{-1}(I + a^\mu \delta_l^2) \mu_l \\ D_l^\delta = D_l^\mu = I + b^\delta \delta_l^2 + c^\delta \delta_l^4, \quad l = 1, 2$$

where a^δ , a^μ , b^δ and c^δ are the parameters (10)–(12) giving the order $2p - 2$.

Then we replace $\frac{\partial f_2}{\partial x_2}$ by a standard Padé approximation $(\frac{\nabla'_2}{\delta x_2})_{j_1, j_2}$ with

$$\nabla'_l = (D'_l)^{-1}(I + a' \delta_l^2) \delta_l \mu_l, \quad D'_l = I + b' \delta_l^2 + c' \delta_l^4, \quad l = 1, 2 \quad (20)$$

where a' , b' and c' are the parameters giving the order $2p - 2$.

Finally, we apply the operator $D_1^\delta D_2'$ to all the terms and obtain

$$(\tilde{r}_1)_{j_1+\frac{1}{2}, j_2} = \left\{ D_2' \left[(I + a^\mu \delta_1^2) \mu_1 \frac{\partial w}{\partial t} + (I + a^\delta \delta_1^2) \frac{\delta_1 f_1}{\delta x_1} \right] + (I + a^\mu \delta_1^2)(I + a' \delta_2^2) \frac{\delta_2 \mu_2 \mu_1 f_2}{\delta x_2} \right\}_{j_1+\frac{1}{2}, j_2}$$

Note that the operator D_l^δ has completely disappeared and thus the parameters b^δ and c^δ play no role. Similarly we construct the second residual in dissipation:

$$(\tilde{r}_2)_{j_1, j_2+\frac{1}{2}} = \left\{ D_1' \left[(I + a^\mu \delta_2^2) \mu_2 \frac{\partial w}{\partial t} + (I + a^\delta \delta_2^2) \frac{\delta_2 f_2}{\delta x_2} \right] + (I + a^\mu \delta_2^2)(I + a' \delta_1^2) \frac{\delta_1 \mu_1 \mu_2 f_1}{\delta x_1} \right\}_{j_1, j_2+\frac{1}{2}}$$

Thus, the residuals in dissipation depend on the five parameters a^δ , a^μ , a' , b' and c' . The parameters of the RBC_3 -dissipation are all zero, those of the RBC_5 -dissipation satisfy (11) and (4) and those of the RBC_7 -dissipation satisfy (12) and (5).

2.3. RBC dissipation

We now discuss the dissipation term of the RBC scheme:

$$\tilde{d}_{j_1, j_2} = \frac{1}{2}[\delta_1(\Phi_1 \tilde{r}_1) + \delta_2(\Phi_2 \tilde{r}_2)]_{j_1, j_2} \quad (21)$$

The matrices involved are defined as:

$$(\Phi_1)_{j_1+\frac{1}{2}, j_2} = [T_{A_1} \text{Diag}(\phi_1^{(i)}) T_{A_1}^{-1}]_{j_1+\frac{1}{2}, j_2}, \quad (\Phi_2)_{j_1, j_2+\frac{1}{2}} = [T_{A_2} \text{Diag}(\phi_2^{(i)}) T_{A_2}^{-1}]_{j_1, j_2+\frac{1}{2}}$$

where, owing to hyperbolicity of System (15), T_{A_l} is an invertible matrix having the right eigenvectors of the flux Jacobian-matrix A_l ($l = 1, 2$) as column vectors and $\text{Diag}(\phi_l^{(i)})$ denotes a diagonal matrix with diagonal entries:

$$\phi_1^{(i)} = \text{sgn}(a_1^{(i)}) \min \left(1, \frac{\delta x_2 |a_1^{(i)}|}{\delta x_1 m(A_2)} \right), \quad \phi_2^{(i)} = \text{sgn}(a_2^{(i)}) \min \left(1, \frac{\delta x_1 |a_2^{(i)}|}{\delta x_2 m(A_1)} \right)$$

where $a_l^{(i)}$ is an eigenvalue of A_l and $m(A_l) = \min_i |a_l^{(i)}|$, $l = 1, 2$. Eigenvalues and eigenvectors of A_1 at $(j_1 + \frac{1}{2}, j_2)$ and of A_2 at $(j_1, j_2 + \frac{1}{2})$, i.e. on cell faces, are computed using the Roe average [22].

The matrices Φ_1 and Φ_2 have been designed in [23,7] to introduce some kind of upwinding. For a one-dimensional problem, Φ_1 reduces to the sign matrix:

$$(\Phi_1)_{j_1 + \frac{1}{2}, j_2} = [\text{sgn}(A_1)]_{j_1 + \frac{1}{2}, j_2} = \left[T_{A_1} \text{Diag}(\text{sgn}(a_1^{(i)})) T_{A_1}^{-1} \right]_{j_1 + \frac{1}{2}, j_2} \quad (22)$$

For a two-dimensional scalar equation:

$$\frac{\partial w}{\partial t} + A_1 \frac{\partial w}{\partial x_1} + A_2 \frac{\partial w}{\partial x_2} = 0 \quad (23)$$

with $A_l = A_l(w) \neq 0$, $l = 1, 2$, Φ_1 and Φ_2 reduces to

$$(\Phi_1)_{j_1 + \frac{1}{2}, j_2} = \left[\text{sgn}(A_1) \min(1, \frac{1}{\theta}) \right]_{j_1 + \frac{1}{2}, j_2}, \quad (\Phi_2)_{j_1, j_2 + \frac{1}{2}} = [\text{sgn}(A_2) \min(1, \theta)]_{j_1, j_2 + \frac{1}{2}} \quad (24)$$

where θ characterizes the advection direction with respect to the mesh:

$$\theta = \frac{\delta x_1 |A_2|}{\delta x_2 |A_1|}$$

For $\theta = 1$, the advection is along a mesh diagonal. For $\theta < 1$, the advection is closer to the x_1 -direction than to the x_2 -direction. For $\theta > 1$, it is the opposite.

The dissipation term (21) has been analyzed in [11] for unsteady problems. For steps δx_1 and δx_2 of the same order of magnitude, say $\mathcal{O}(h)$, a Taylor expansion of (21) for a RBC scheme of order $2p - 1$ gives:

$$\begin{aligned} \tilde{d}_{j_1, j_2} = & (-1)^{p-1} \kappa \left\{ \frac{\partial}{\partial x_1} \left[\Phi_1 (\delta x_1^{2p-1} \frac{\partial^{2p-1} f_1}{\partial x_1^{2p-1}} + \chi \delta x_1 \delta x_2^{2p-2} \frac{\partial^{2p-1} f_2}{\partial x_2^{2p-1}}) \right] \right. \\ & \left. + \frac{\partial}{\partial x_2} \left[\Phi_2 (\chi \delta x_1^{2p-2} \delta x_2 \frac{\partial^{2p-1} f_1}{\partial x_1^{2p-1}} + \delta x_2^{2p-1} \frac{\partial^{2p-1} f_2}{\partial x_2^{2p-1}}) \right] \right\}_{j_1, j_2} + \mathcal{O}(h^{2p+1}) \end{aligned} \quad (25)$$

where $\kappa > 0$ and χ are two constant coefficients depending only on the parameters in dissipation. Note that (25) does not contain time derivatives since they have been replaced by space derivatives using the exact system (15).

For the RBC₃-dissipation:

$$\kappa = \frac{1}{24}, \quad \chi = 12 \left(b' - \frac{1}{6} \right)$$

For the RBC₅-dissipation:

$$\kappa = \frac{1}{240}, \quad \chi = 20 \left(b' - \frac{1}{5} \right)$$

For the RBC₇-dissipation:

$$\kappa = \frac{1}{5600}, \quad \chi = \frac{280}{3} \left(c' - \frac{1}{70} \right)$$

Considering the scalar equation (23), it has been proved [11] that a necessary and sufficient condition for the partial differential operator in (25) be dissipative, for all advection directions (A_1, A_2) and for (Φ_1, Φ_2) of the form (24), is $\chi = 0$, i.e. no crossed derivatives in (25). A similar result has also been proved in 3-D [11].

For the RBC₇-dissipation, $\chi = 0$ means $c' = \frac{1}{70}$. Together with the condition

$$a' = \frac{1}{30} + 6c', \quad b' = \frac{1}{5} + 6c'$$

ensuring the 6th-order accuracy of $\frac{\nabla'}{\delta x_2}$ in \tilde{r}_1 and $\frac{\nabla'}{\delta x_1}$ in \tilde{r}_2 , we obtain:

$$a' = \frac{5}{42}, \quad b' = \frac{2}{7}, \quad c' = \frac{1}{70}$$

that is a 8th-order accuracy on these ∇' -Padé derivatives as for the ∇ -Padé derivatives in the main residual. A similar situation occurs with RBC_5 - and RBC_3 -dissipation. Generally speaking, a correct dissipation requires $\nabla' = \nabla$, that is

$$a' = a, \quad b' = b, \quad c' = c$$

Fortunately, this accuracy enhancement of some terms in the dissipation operator does not extend the scheme stencil.

Finally, the dissipative RBC schemes on a 5×5 stencil are defined by the main residual (19) and the residuals in dissipation:

$$(\tilde{r}_1)_{j_1+\frac{1}{2}, j_2} = \left\{ D_2 \left[(I + a^\mu \delta_1^2) \mu_1 \frac{\partial w}{\partial t} + (I + a^\delta \delta_1^2) \frac{\delta_1 f_1}{\delta x_1} \right] + (I + a^\mu \delta_1^2)(I + a \delta_2^2) \frac{\delta_2 \mu_2 \mu_1 f_2}{\delta x_2} \right\}_{j_1+\frac{1}{2}, j_2} \quad (26)$$

$$(\tilde{r}_2)_{j_1, j_2+\frac{1}{2}} = \left\{ D_1 \left[(I + a^\mu \delta_2^2) \mu_2 \frac{\partial w}{\partial t} + (I + a^\delta \delta_2^2) \frac{\delta_2 f_2}{\delta x_2} \right] + (I + a^\mu \delta_2^2)(I + a \delta_1^2) \frac{\delta_1 \mu_1 \mu_2 f_1}{\delta x_1} \right\}_{j_1, j_2+\frac{1}{2}} \quad (27)$$

with the following parameters:

$$RBC_3: \quad a = 0, \quad b = \frac{1}{6}, \quad c = a^\mu = a^\delta = 0 \quad (28)$$

$$RBC_5: \quad a = \frac{1}{30}, \quad b = \frac{1}{5}, \quad c = a^\mu = 0, \quad a^\delta = \frac{1}{12} \quad (29)$$

$$RBC_7: \quad a = \frac{5}{42}, \quad b = \frac{2}{7}, \quad c = \frac{1}{70}, \quad a^\mu = \frac{1}{10}, \quad a^\delta = \frac{11}{60} \quad (30)$$

In these schemes, the dissipation error (of order $2p - 1$) dominates the dispersive error (of order $2p$), which is a good feature for robustness. Spectral properties of these schemes are described in [12].

Since the RBC_5 scheme is mostly used in our numerical applications, we write down its specific residuals (omitting the subscripts):

$$\begin{aligned} \tilde{r} &= (I + \frac{1}{5} \delta_1^2)(I + \frac{1}{5} \delta_2^2) \frac{\partial w}{\partial t} + (I + \frac{1}{5} \delta_2^2)(I + \frac{1}{30} \delta_1^2) \frac{\delta_1 \mu_1 f_1}{\delta x_1} + (I + \frac{1}{5} \delta_1^2)(I + \frac{1}{30} \delta_2^2) \frac{\delta_2 \mu_2 f_2}{\delta x_2} \\ \tilde{r}_1 &= (I + \frac{1}{5} \delta_2^2) \left[\mu_1 \frac{\partial w}{\partial t} + (I + \frac{1}{12} \delta_1^2) \frac{\delta_1 f_1}{\delta x_1} \right] + (I + \frac{1}{30} \delta_2^2) \frac{\delta_2 \mu_2 \mu_1 f_2}{\delta x_2} \\ \tilde{r}_2 &= (I + \frac{1}{5} \delta_1^2) \left[\mu_2 \frac{\partial w}{\partial t} + (I + \frac{1}{12} \delta_2^2) \frac{\delta_2 f_2}{\delta x_2} \right] + (I + \frac{1}{30} \delta_1^2) \frac{\delta_1 \mu_1 \mu_2 f_1}{\delta x_1} \end{aligned}$$

The RBC_5 -dissipation approximates:

$$\tilde{d} = \frac{1}{240} \left[\delta x_1^5 \frac{\partial}{\partial x_1} (\Phi_1 \frac{\partial^5 f_1}{\partial x_1^5}) + \delta x_2^5 \frac{\partial}{\partial x_2} (\Phi_2 \frac{\partial^5 f_2}{\partial x_2^5}) \right] + \mathcal{O}(\delta x_1^7, \delta x_2^7)$$

2.4. Direct extraction of the time-derivative

A difficulty inherent to the RBC schemes is that the time derivative occurs, through linear discrete operators due to compactness, not only in the main residual but also in the other two residuals (or three residuals in 3-D) involved in the numerical dissipation. To overcome this difficulty, the following approach has been proposed in [13]. First, the main residual and the dissipation are split into a part containing the time derivative and a purely-spatial part (still omitting the subscripts j_1, j_2):

$$\tilde{r} = D_1 D_2 \frac{\partial w}{\partial t} + \tilde{r}^0$$

$$\tilde{d} = \frac{1}{2} \left\{ \delta_1 \left[\Phi_1 D_2 (I + a^\mu \delta_1^2) \mu_1 \cdot \right] + \delta_2 \left[\Phi_2 D_1 (I + a^\mu \delta_2^2) \mu_2 \cdot \right] \right\} \frac{\partial w}{\partial t} + \frac{1}{2} [\delta_1 (\Phi_1 \tilde{r}_1^0) + \delta_2 (\Phi_2 \tilde{r}_2^0)]$$

where

$$\begin{aligned}
\tilde{r}^0 &= D_2(I + a\delta_1^2) \frac{\delta_1 \mu_1 f_1}{\delta x_1} + D_1(I + a\delta_2^2) \frac{\delta_2 \mu_2 f_2}{\delta x_2} \\
\tilde{r}_1^0 &= D_2(I + a^\delta \delta_1^2) \frac{\delta_1 f_1}{\delta x_1} + (I + a^\mu \delta_1^2)(I + a\delta_2^2) \frac{\delta_2 \mu_2 \mu_1 f_2}{\delta x_2} \\
\tilde{r}_2^0 &= D_1(I + a^\delta \delta_2^2) \frac{\delta_2 f_2}{\delta x_2} + (I + a^\mu \delta_2^2)(I + a\delta_1^2) \frac{\delta_1 \mu_1 \mu_2 f_1}{\delta x_1}
\end{aligned}$$

Since the discrete operators δ_1 , δ_2 , μ_1 and μ_2 commute two by two, we can also write \tilde{d} as

$$\tilde{d} = (M_1 D_2 + M_2 D_1) \frac{\partial w}{\partial t} + \frac{1}{2} [\delta_1 (\Phi_1 \tilde{r}_1^0) + \delta_2 (\Phi_2 \tilde{r}_2^0)]$$

with

$$M_1 = \frac{1}{2} \delta_1 \left[\Phi_1 (I + a^\mu \delta_1^2) \mu_1 \right], \quad M_2 = \frac{1}{2} \delta_2 \left[\Phi_2 (I + a^\mu \delta_2^2) \mu_2 \right]$$

Owing to the matrix-coefficients Φ_1 and Φ_2 , the matrices associated to the discrete operators M_1 and M_2 have a block structure. They are block-tridiagonal if $a^\mu = 0$, i.e. for RBC_3 and RBC_5 , or block-pentadiagonal for RBC_7 . Their dimension is the number of mesh points in x_1 - or x_2 -direction with a block-size equal to the number of equations in the exact system (15). For consistency, the same block-structure is given to the matrices associated to D_1 and D_2 with diagonal blocks. The matrices associated to D_1 and D_2 are block-tridiagonal if $c = 0$, i.e. for RBC_3 and RBC_5 , or block-pentadiagonal for RBC_7 .

The RBC scheme $\tilde{r} = \tilde{d}$ can thus be rewritten as

$$\Lambda \frac{\partial w}{\partial t} = H_0 \quad (31)$$

where Λ denotes the discrete operator:

$$\Lambda = D_1 D_2 - M_1 D_2 - M_2 D_1$$

and H_0 is the space-flux contribution:

$$H_0 = -\tilde{r}^0 + \frac{1}{2} [\delta_1 (\Phi_1 \tilde{r}_1^0) + \delta_2 (\Phi_2 \tilde{r}_2^0)] \quad (32)$$

The time derivative can be obtained by solving the linear system (31). However, the matrix associated to Λ is rather complicated. To simplify it, a dimensional factorization of Λ is attempted as:

$$\Lambda = \Lambda_1 \Lambda_2 - \Lambda_c$$

where

$$\Lambda_1 = D_1 - M_1, \quad \Lambda_2 = D_2 - M_2, \quad \Lambda_c = M_1 M_2$$

As D_1 and D_2 , the one-dimensional operators Λ_1 and Λ_2 are locally $\mathcal{O}(1)$, but since $M_1 = \mathcal{O}(\delta x_1)$ and $M_2 = \mathcal{O}(\delta x_2)$, the corrective operator Λ_c is locally $\mathcal{O}(\delta x_1 \delta x_2)$ and cannot be neglected. Therefore the linear system (31) is considered in the form:

$$\Lambda_1 \Lambda_2 \frac{\partial w}{\partial t} = H_0 + \Lambda_c \frac{\partial w}{\partial t} \quad (33)$$

and solved approximately in a few iterations as:

$$\Lambda_1 \Lambda_2 \frac{\partial w^{(m+1)}}{\partial t} = H_0 + \Lambda_c \frac{\partial w^{(m)}}{\partial t}, \quad m = 0, 1, \dots, m_f \quad (34)$$

starting from

$$\frac{\partial w^{(0)}}{\partial t} = 0. \quad (35)$$

Since Eq. (34) can be split into

$$\begin{cases} \Lambda_1 \frac{\partial \tilde{w}}{\partial t} = H_0 + \Lambda_c \frac{\partial w^{(m)}}{\partial t} \\ \Lambda_2 \frac{\partial w^{(m+1)}}{\partial t} = \frac{\partial \tilde{w}}{\partial t} \end{cases} \quad (36)$$

we have only to solve linear systems that are block-tridiagonal for RBC_3 and RBC_5 or block-pentadiagonal for RBC_7 . It has been proved in [13] that one iteration ($m_f = 1$) for RBC_3 , two iterations for RBC_5 and three iterations for RBC_7 are sufficient to preserve the space accuracy order.

The numerical solution is advanced by solving the ordinary differential equation (ODE):

$$\frac{\partial w}{\partial t} = F(w) \quad (37)$$

where $F = \Lambda^{-1}H_0$. Clearly, the main cost per time-step comes from the calculation of F as the solution of the problem (36). So we are interested in ODE methods requiring few evaluations of F per time-step. Several explicit methods of order 4 were compared in [13], namely the Adams–Bashforth method AB_4 , the extrapolated Backward Differential Formula $eBDF_4$, the classical Runge–Kutta method RK_4 and an explicit combination of the Adams–Bashforth and Adams–Moulton methods called ABM_4 . The latter was found to be the least consuming in CPU-time. This ABM_4 method writes:

$$\begin{cases} \tilde{w}^{n+1} = w^n + \frac{\Delta t}{12}(23F^n - 16F^{n-1} + 5F^{n-2}) \\ w^{n+1} = w^n + \frac{\Delta t}{24}(9\tilde{F}^{n+1} + 19F^n - 5F^{n-1} + F^{n-2}) \end{cases} \quad (38)$$

where $w^n = (w_{j_1, j_2}^n)$ denotes the numerical solution at time level $t^n = n\Delta t$, $F^n = F(w^n)$ and $\tilde{F}^{n+1} = F(\tilde{w}^{n+1})$.

The maximal time-step for stability of ABM_4 is lower than that of RK_4 but much greater than those of $eBDF_4$ and AB_4 . On the other hand, ABM_4 requires two evaluations of F per time step (F^n and \tilde{F}^{n+1}) while RK_4 requires four. Finally, the shortest computing-time for the RBC schemes was obtained with ABM_4 .

3. A new compact formulation

In order to simplify the RBC scheme (17), the idea is to compute the time derivative in dissipation from the non-dissipative scheme

$$\tilde{r} = 0$$

From the definition (19) of the main residual \tilde{r} , we obtain

$$\frac{\partial \tilde{w}}{\partial t} = -\frac{\nabla_1 f_1}{\delta x_1} - \frac{\nabla_2 f_2}{\delta x_2} \quad (39)$$

using the Padé compact derivatives defined by (18). The residuals in dissipation \tilde{r}_1 and \tilde{r}_2 become

$$\begin{aligned} \check{r}_1 &= (D_2)^{-1}\tilde{r}_1 = (I + a^\mu \delta_1^2)\mu_1\left(\frac{\partial \tilde{w}}{\partial t} + \frac{\nabla_2 f_2}{\delta x_2}\right) + (I + a^\delta \delta_1^2)\frac{\delta_1 f_1}{\delta x_1} \\ \check{r}_2 &= (D_1)^{-1}\tilde{r}_2 = (I + a^\mu \delta_2^2)\mu_2\left(\frac{\partial \tilde{w}}{\partial t} + \frac{\nabla_1 f_1}{\delta x_1}\right) + (I + a^\delta \delta_2^2)\frac{\delta_2 f_2}{\delta x_2} \end{aligned}$$

Since \tilde{r}_1, \tilde{r}_2 approximate 0 at a high order and $(D_2)^{-1} = I + \mathcal{O}(\delta x_2^2)$, $(D_1)^{-1} = I + \mathcal{O}(\delta x_1^2)$, the new residuals \check{r}_1, \check{r}_2 approximate 0 at the same order as \tilde{r}_1, \tilde{r}_2 .

Using (39) to eliminate $\partial \tilde{w}/\partial t$, the new residuals in dissipation reduce to the simple expressions:

$$\begin{aligned} \check{r}_1 &= (I + a^\delta \delta_1^2)\frac{\delta_1 f_1}{\delta x_1} - (I + a^\mu \delta_1^2)\mu_1\frac{\nabla_1 f_1}{\delta x_1} \\ \check{r}_2 &= (I + a^\delta \delta_2^2)\frac{\delta_2 f_2}{\delta x_2} - (I + a^\mu \delta_2^2)\mu_2\frac{\nabla_2 f_2}{\delta x_2} \end{aligned} \quad (40)$$

and the new compact scheme takes the form:

$$\frac{\partial w}{\partial t} + \frac{\nabla_1 f_1}{\delta x_1} + \frac{\nabla_2 f_2}{\delta x_2} = \frac{1}{2}[\delta_1(\Phi_1 \check{r}_1) + \delta_2(\Phi_2 \check{r}_2)] \quad (41)$$

For the parameter sets (28), (29) and (30), this scheme remains respectively accurate at order 3, 5 and 7, but it is much simpler to implement:

- the scheme (41) only requires the solution of algebraic linear systems with constant coefficients to compute the Padé derivatives $\nabla_1 f_1/\delta x_1$ and $\nabla_2 f_2/\delta x_2$ and these systems are tridiagonal for scheme-order 3 and 5 or pentadiagonal for order 7, without any block.
- Provided the above Padé derivatives are stored, the calculation of the residuals in dissipation requires no system inversion and is fast, specially for order 5 ($a^\mu = 0$) and of course for order 3 ($a^\delta = a^\mu = 0$).

– Furthermore, the new scheme is stable with simpler Φ_1 and Φ_2 matrices than those recalled in Section 2.3 for the original RBC scheme. All the present computations with the new scheme have been run using:

$$\Phi_l = \text{sgn}(A_l) = T_{A_l} \text{Diag}[\text{sgn}(a_l^{(i)})] (T_{A_l})^{-1}, \quad l = 1, 2 \quad (42)$$

The linear stability analysis presented below for the new scheme with (42) justifies this simplification for 2-D and 3-D scalar problems.

Although this is not recommended for an efficient programming, we now develop the dissipation residuals (40) in order to precisely identify the new form of dissipation. From the definition (18) of ∇_l :

$$D_l \mu_l \frac{\nabla_l f_l}{\delta x_l} = \mu_l D_l \frac{\nabla_l f_l}{\delta x_l} = \mu_l^2 (I + a \delta_l^2) \frac{\delta_l f_l}{\delta x_l}, \quad l = 1, 2$$

Using the relation $\mu_l^2 = I + \frac{1}{4} \delta_l^2$, we obtain:

$$D_l \mu_l \frac{\nabla_l f_l}{\delta x_l} = (I + \frac{1}{4} \delta_l^2) (I + a \delta_l^2) \frac{\delta_l f_l}{\delta x_l}, \quad l = 1, 2$$

The residuals (40) can thus be written as:

$$\check{r}_l = (D_l)^{-1} N_l \frac{\delta_l f_l}{\delta x_l}, \quad l = 1, 2$$

where

$$N_l = D_l (I + a^\delta \delta_l^2) - (I + a^\mu \delta_l^2) (I + \frac{1}{4} \delta_l^2) (I + a \delta_l^2), \quad D_l = I + b \delta_l^2 + c \delta_l^4$$

A simple calculation reduces N_l to

$$N_l = n_2 \delta_l^2 + n_4 \delta_l^4 + n_6 \delta_l^6$$

with

$$n_2 = b - a + a^\delta - a^\mu - \frac{1}{4}, \quad n_4 = b a^\delta - a a^\mu + c - \frac{a}{4} - \frac{a^\mu}{4}, \quad n_6 = c a^\delta - \frac{a a^\mu}{4}$$

For the parameters (28) of RBC_3 : $n_2 = -\frac{1}{12}$, $n_4 = 0$, $n_6 = 0$

For the parameters (29) of RBC_5 : $n_2 = 0$, $n_4 = \frac{1}{120}$, $n_6 = 0$

For the parameters (30) of RBC_7 : $n_2 = 0$, $n_4 = 0$, $n_6 = -\frac{1}{2800}$

Inserting

$$\check{r}_l = (I + b \delta_l^2 + c \delta_l^4)^{-1} (n_2 \delta_l^2 + n_4 \delta_l^4 + n_6 \delta_l^6) \frac{\delta_l f_l}{\delta x_l}, \quad l = 1, 2$$

in the right-hand side of (41) and replacing ∇_l by (18) in the left-hand side, the new compact scheme takes the form:

$$\text{at order 3: } \frac{\partial w}{\partial t} + \sum_{l=1}^2 (I + \frac{1}{6} \delta_l^2)^{-1} \frac{\delta_l \mu_l f_l}{\delta x_l} = -\frac{1}{24} \sum_{l=1}^2 \frac{\delta_l}{\delta x_l} \left[\text{sgn}(A_l) (I + \frac{1}{6} \delta_l^2)^{-1} \delta_l^3 f_l \right] \quad (43)$$

$$\text{at order 5: } \frac{\partial w}{\partial t} + \sum_{l=1}^2 (I + \frac{1}{5} \delta_l^2)^{-1} (I + \frac{1}{30} \delta_l^2) \frac{\delta_l \mu_l f_l}{\delta x_l} = \frac{1}{240} \sum_{l=1}^2 \frac{\delta_l}{\delta x_l} \left[\text{sgn}(A_l) (I + \frac{1}{5} \delta_l^2)^{-1} \delta_l^5 f_l \right] \quad (44)$$

$$\text{at order 7: } \frac{\partial w}{\partial t} + \sum_{l=1}^2 (I + \frac{2}{7} \delta_l^2 + \frac{1}{70} \delta_l^4)^{-1} (I + \frac{5}{42} \delta_l^2) \frac{\delta_l \mu_l f_l}{\delta x_l} = -\frac{1}{5600} \sum_{l=1}^2 \frac{\delta_l}{\delta x_l} \left[\text{sgn}(A_l) (I + \frac{2}{7} \delta_l^2 + \frac{1}{70} \delta_l^4)^{-1} \delta_l^7 f_l \right] \quad (45)$$

Clearly, the dissipation term of each above scheme expands as in (25) with $\chi = 0$ and the same coefficient κ . It therefore behaves in first approximation as the dissipation of the basic RBC scheme. Note that the new scheme remains compact in its main part (left-hand side), but not in its dissipative part. In practice, this compacity loss is not too serious, because it is easier to locally modify the dissipation than the consistent part (for instance near a boundary) and also, as we see below, this dissipation is only used once per time step, at the last stage of a Runge-Kutta method.

Let us now compare the new scheme to a classical upwind-scheme. The latter can be written with a centered non-compact approximation plus a numerical dissipation as follows (see [9] for instance):

$$\text{at order 3: } \frac{\partial w}{\partial t} + \sum_{l=1}^2 \left(I - \frac{1}{6} \delta_l^2 \right) \frac{\delta_l \mu_l f_l}{\delta x_l} = -\frac{1}{12} \sum_{l=1}^2 \frac{\delta_l}{\delta x_l} \left(|A_l| \delta_l^3 w \right) \quad (46)$$

$$\text{at order 5: } \frac{\partial w}{\partial t} + \sum_{l=1}^2 \left(I - \frac{1}{6} \delta_l^2 + \frac{1}{30} \delta_l^4 \right) \frac{\delta_l \mu_l f_l}{\delta x_l} = \frac{1}{60} \sum_{l=1}^2 \frac{\delta_l}{\delta x_l} \left(|A_l| \delta_l^5 w \right) \quad (47)$$

$$\text{at order 7: } \frac{\partial w}{\partial t} + \sum_{l=1}^2 \left(I - \frac{1}{6} \delta_l^2 + \frac{1}{30} \delta_l^4 - \frac{1}{140} \delta_l^6 \right) \frac{\delta_l \mu_l f_l}{\delta x_l} = -\frac{1}{280} \sum_{l=1}^2 \frac{\delta_l}{\delta x_l} \left(|A_l| \delta_l^7 w \right) \quad (48)$$

where

$$|A_l| = T_{A_l} \text{Diag}[(a_l^{(i)})] (T_{A_l})^{-1}, \quad l = 1, 2$$

Note that the dissipation coefficient of the upwind scheme is always half of the last coefficient in the left-hand side (see [9] for a proof).

Comparing centered approximations of a first derivative in Section 2.1, we have mentioned that a compact formula is more accurate than a non-compact one of the same accuracy order. This yields that the consistent part (left-hand-side) of the new schemes (43)–(45) produces less dispersive-error than the consistent part of the upwind schemes (46)–(48). Now concerning the dissipation, there is some analogy between the right-hand side of the two types of high-order schemes. For a linear problem (constant matrices A_1 and A_2), these two scheme-types have the same form of first differential-approximation, but the schemes (43)–(45) are less dissipative because their dissipation coefficients are smaller, namely 2, 4 and 20 times smaller at order 3, 5 and 7, respectively. So we can conclude that both dispersive and dissipative errors are smaller for the schemes (43)–(45).

We return to the usual form (40)–(41) of the new schemes and present their time integration. Due to the simplicity of these schemes, the most efficient time-integration is no longer ABM_4 as in Section 2.4 for the basic RBC -schemes, but has been found in the class of Runge–Kutta methods. Specifically, we use the low-storage explicit Runge–Kutta method $RK06$ proposed by Bogey and Bailly [21]. This six-stage method was constructed by optimizing its dispersion and dissipation properties. It has been shown [21] to be more stable and more accurate than the standard low-storage method with four stages ($RKs4$).

For a differential equation of the form (37), the $RK06$ method reads:

$$\begin{aligned} w^{(0)} &= w^n \\ w^{(k)} &= w^n + \alpha_k \Delta t F(w^{(k-1)}), \quad k = 1, 2, \dots, 6 \\ w^{n+1} &= w^{(6)} \end{aligned} \quad (49)$$

with the following coefficients:

$$\begin{aligned} \alpha_1 &= 0.117979901657, \quad \alpha_2 = 0.184646966491, \quad \alpha_3 = 0.246623604310, \quad \alpha_4 = 0.331839542736, \\ \alpha_5 &= \frac{1}{2}, \quad \alpha_6 = 1. \end{aligned}$$

More precisely, the time integration of the new scheme will be done by taking into account the numerical dissipation at the last stage only. This reduces the computing time but not the accuracy order because the numerical dissipation is of the order of the scheme spatial-accuracy (3, 5 or 7). So we replace (49) by

$$w^{(k)} = w^n + \alpha_k \Delta t F_k(w^{(k-1)}), \quad k = 1, 2, \dots, 6 \quad (50)$$

with

$$F_k = -\frac{\nabla_1 f_1}{\delta x_1} - \frac{\nabla_2 f_2}{\delta x_2} + \frac{\chi_k}{2} [\delta_1(\Phi_1 \check{r}_1) + \delta_2(\Phi_2 \check{r}_2)] \quad (51)$$

where $\chi_k = 0$ for $k = 1, 2, \dots, 5$. The coefficient χ_6 is normally equal to 1, but it may be reduced in smooth flows.

We now present a L_2 -stability analysis of the method (50)–(51) for a 2-D advection problem ($f_1 = A_1 w$ and $f_2 = A_2 w$ where A_1 and A_2 are scalar constants). The Fourier symbol (in space) of the Padé derivative $\Delta t A_l \nabla_l / \delta x_l$ is $\iota \dot{A}_l P_l$, with $\dot{A}_l = \Delta t A_l / \delta x_l$ and

$$P_l = \frac{[1 - 2a(1 - \cos \xi_l)] \sin \xi_l}{1 - 2b(1 - \cos \xi_l) + 4c(1 - \cos \xi_l)^2}, \quad l = 1, 2$$

where ξ_l is the reduced wave-number in the x_l -direction.

Concerning the dissipation, the Fourier symbol of $\frac{1}{2} \Delta t \delta_l(\Phi_l \check{r}_l)$ is $-|\dot{A}_l| Q_l$, where

$$Q_l = [1 - 2a^\delta(1 - \cos \xi_l)](1 - \cos \xi_l) - \frac{1}{2}[1 - 2a^\mu(1 - \cos \xi_l)] \sin \xi_l P_l, \quad l = 1, 2$$

The Fourier symbol of $\Delta t F_k$ is therefore

$$\Delta t \hat{F}_k = -(Q^{(k)} + \iota P)$$

where $P = \dot{A}_1 P_1 + \dot{A}_2 P_2$ and $Q^{(k)} = \chi_k(|\dot{A}_1| Q_1 + |\dot{A}_2| Q_2)$.

The amplification factor $g = g(\xi_1, \xi_2)$ governing the Fourier-transform evolution of w^n ($\hat{w}^{n+1} = g \hat{w}^n$) satisfies

$$\begin{aligned} g^{(0)} &= 1 \\ g^{(k)} &= 1 - \alpha_k(Q^{(k)} + \iota P) g^{(k-1)}, \quad k = 1, 2, \dots, 6 \\ g &= g^{(6)} \end{aligned}$$

Denoting $g_r^{(k)} = \Re(g^{(k)})$ and $g_i^{(k)} = \Im(g^{(k)})$, we obtain:

$$\begin{aligned} g_r^{(0)} &= 1, \quad g_i^{(0)} = 0 \\ g_r^{(k)} &= 1 - \alpha_k(Q^{(k)} g_r^{(k-1)} - P g_i^{(k-1)}), \quad g_i^{(k)} = -\alpha_k(P g_r^{(k-1)} + Q^{(k)} g_i^{(k-1)}), \quad k = 1, 2, \dots, 6 \\ |g|^2 &= (g_r^{(6)})^2 + (g_i^{(6)})^2 \end{aligned}$$

The stability domain of the scheme (50)–(51), that is the part of the (\dot{A}_1, \dot{A}_2) -plane in which $|g(\xi_1, \xi_2)| \leq 1$ for all (ξ_1, ξ_2) , is plotted on Fig. 1 for various space-orders corresponding to scheme coefficients (28)–(30) and for various dissipation-coefficients χ_6 in the last RK6-stage. The stability condition is of the form:

$$\Delta t \left(\frac{|A_1|}{\delta x_1} + \frac{|A_2|}{\delta x_2} \right) \leq \eta \quad (52)$$

The stability domain increases as the space accuracy increases or as the dissipation coefficient decreases. With $\chi_6 = 1$, the η -bound is equal to 1, 1.30 and 1.80 for the space-orders 3, 5 and 7, respectively. For a 5th-order accuracy in space, $\eta = 1.30, 1.82$ and 1.98 for $\chi_6 = 1, 0.5$ and 0.2 , respectively. Note that in a square mesh ($\delta x_1 = \delta x_2 = \delta x$), a sufficient condition for stability is

$$\frac{\Delta t}{\delta x} \sqrt{A_1^2 + A_2^2} \leq \frac{\eta}{\sqrt{2}}$$

Consider now the three-dimensional hyperbolic system

$$\frac{\partial w}{\partial t} + \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3} = 0 \quad (53)$$

with $f_3 = f_3(w)$ and $A_3 = df_3/dw$.

The extension of the new compact scheme to (53) is really straightforward. The time integration (50) has simply to be applied to

$$F_k = -\frac{\nabla_1 f_1}{\delta x_1} - \frac{\nabla_2 f_2}{\delta x_2} - \frac{\nabla_3 f_3}{\delta x_3} + \frac{\chi_k}{2} [\delta_1(\Phi_1 \check{r}_1) + \delta_2(\Phi_2 \check{r}_2) + \delta_3(\Phi_3 \check{r}_3)] \quad (54)$$

where $\chi_k = 0$ for $k = 1, 2, \dots, 5$. The third Padé derivative $\nabla_3 f_3 / \delta x_3$ is defined similarly as in (18) and the third residual in dissipation \check{r}_3 as in (40) with a matrix Φ_3 given by (42) for $l = 3$.

The 3-D linear stability analysis is quite similar to the 2-D one and produces a stability domain within a regular octahedron in the $(\dot{A}_1, \dot{A}_2, \dot{A}_3)$ -space. The boundary points of this octahedron on the three axis are $x_l = \pm \eta$, $l = 1, 2, 3$. The 3-D stability condition is thus of the form:

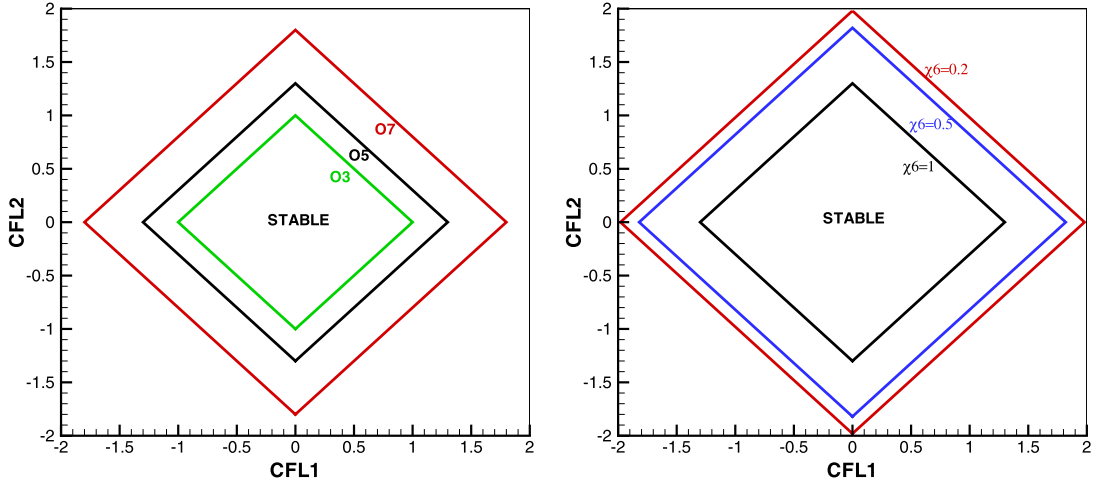


Fig. 1. 2D-stability domain of the compact scheme (50)–(51) for an advection problem in terms of $\text{CFL1} = \Delta t A_1 / \delta x_1$ and $\text{CFL2} = \Delta t A_2 / \delta x_2$. Left: Space-order 3, 5, 7 and dissipation-coefficient $\chi_6 = 1$. Right: Space-order 5 and $\chi_6 = 0.2, 0.5, 1$.

$$\Delta t \left(\frac{|A_1|}{\delta x_1} + \frac{|A_2|}{\delta x_2} + \frac{|A_3|}{\delta x_3} \right) \leq \eta \quad (55)$$

with the same η -values as those given above for the 2-D case.

4. The new compact scheme for the 3-D compressible Navier–Stokes equations

We present the extension of the compact scheme (50), (54) at order 5 to the compressible Navier–Stokes equations. In dimensionless form, these equations can be written with Cartesian coordinates as

$$\frac{\partial w}{\partial t} + \frac{\partial}{\partial x_1}(f_1^E - f_1^V) + \frac{\partial}{\partial x_2}(f_2^E - f_2^V) + \frac{\partial}{\partial x_3}(f_3^E - f_3^V) = 0 \quad (56)$$

with the conservative variables:

$$w = [\rho, \rho u_1, \rho u_2, \rho u_3, \rho E]^T$$

and the following Euler and viscous fluxes:

$$\begin{aligned} f_1^E &= [\rho u_1, \rho u_1^2 + p, \rho u_1 u_2, \rho u_1 u_3, \rho u_1 H]^T & f_1^V &= \frac{1}{Re} [0, \tau_{11}, \tau_{12}, \tau_{13}, u_1 \tau_{11} + u_2 \tau_{12} + u_3 \tau_{13} - q_1]^T \\ f_2^E &= [\rho u_2, \rho u_2 u_1, \rho u_2^2 + p, \rho u_2 u_3, \rho u_2 H]^T & f_2^V &= \frac{1}{Re} [0, \tau_{21}, \tau_{22}, \tau_{23}, u_1 \tau_{21} + u_2 \tau_{22} + u_3 \tau_{23} - q_2]^T \\ f_3^E &= [\rho u_3, \rho u_3 u_1, \rho u_3 u_2, \rho u_3^2 + p, \rho u_3 H]^T & f_3^V &= \frac{1}{Re} [0, \tau_{31}, \tau_{32}, \tau_{33}, u_1 \tau_{31} + u_2 \tau_{32} + u_3 \tau_{33} - q_3]^T \end{aligned}$$

Assuming that the fluid is Newtonian and satisfies the Stokes relation, the Fourier law for heat conduction and the perfect gas law with a constant specific-heat ratio γ , the components of the shear stress tensor τ and of the heat conduction flux q can be written as:

$$\begin{aligned} \tau_{11} &= \frac{4}{3} \nu \frac{\partial u_1}{\partial x_1} - \frac{2}{3} \nu \left(\frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} \right), & \tau_{22} &= \frac{4}{3} \nu \frac{\partial u_2}{\partial x_2} - \frac{2}{3} \nu \left(\frac{\partial u_3}{\partial x_3} + \frac{\partial u_1}{\partial x_1} \right), & \tau_{33} &= \frac{4}{3} \nu \frac{\partial u_3}{\partial x_3} - \frac{2}{3} \nu \left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} \right) \\ \tau_{12} &= \tau_{21} = \nu \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right), & \tau_{13} &= \tau_{31} = \nu \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right), & \tau_{23} &= \tau_{32} = \nu \left(\frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2} \right) \\ q_1 &= -\frac{\gamma \nu}{Pr} \frac{\partial e}{\partial x_1}, & q_2 &= -\frac{\gamma \nu}{Pr} \frac{\partial e}{\partial x_2}, & q_3 &= -\frac{\gamma \nu}{Pr} \frac{\partial e}{\partial x_3} \end{aligned}$$

In the above formulas, ρ is the density, p the pressure, (u_1, u_2, u_3) are the Cartesian components of the fluid velocity \vec{U} , $e = \frac{1}{\gamma-1} \frac{p}{\rho}$ is the specific internal energy, $E = e + \frac{1}{2}(u_1^2 + u_2^2 + u_3^2)$ the specific total energy, $H = E + \frac{p}{\rho}$ the specific total enthalpy and Re, Pr are the Reynolds and Prandtl numbers. Note that the dynamic-viscosity coefficient has been denoted by ν , rather than by μ as usual, to avoid confusion with the average discrete-operator.

The Euler-flux derivatives are approximated by (54) at order 5, that is using sixth-order Padé derivatives plus the numerical dissipation. Concerning the viscous terms, it is known that a second derivative can also be approximated at order 6 on

a 5-point stencil with a compact formula requiring only the solution of a tridiagonal linear system. This can be done with the Padé formula:

$$\frac{\partial^2 v}{\partial x_l^2} = (I + \frac{2}{15}\delta_l^2)^{-1} (I + \frac{1}{20}\delta_l^2) \frac{\delta_l^2 v}{\delta x_l^2} + \mathcal{O}(\delta x_l^6), \quad l = 1, 2, 3$$

However, in the compressible Navier–Stokes equations, the second differential operators rather appear with variable coefficients. This is true in the energy equation and also in the momentum equations when the viscosity coefficient ν varies with temperature. Actually, we have to discretize terms of the form $\frac{\partial}{\partial x_l} (u \frac{\partial v}{\partial x_l})$ that cannot be approximated at order 6 on a 5-point stencil. This is the reason why we choose to discretize the viscous terms at order 4 and keep a 5-point stencil. Thus, for a second derivative in x_l -direction, we use

$$\frac{\partial}{\partial x_l} (u \frac{\partial v}{\partial x_l}) = \frac{\bar{\nabla}_l}{\delta x_l} \left[(\bar{\mu}_l u) \frac{\bar{\nabla}_l v}{\delta x_l} \right] + \mathcal{O}(\delta x_l^4), \quad l = 1, 2, 3 \quad (57)$$

with 4th-order Padé approximations at midpoints defined by the parameters (13), i.e.

$$\bar{\nabla}_l = (I + \frac{1}{24}\delta_l^2)^{-1} \delta_l, \quad \bar{\mu}_l = (I + \frac{1}{8}\delta_l^2)^{-1} \mu_l$$

For a crossed second-derivative, we use

$$\frac{\partial}{\partial x_l} (u \frac{\partial v}{\partial x_m}) = \frac{\bar{\nabla}_l}{\delta x_l} \left[\bar{\mu}_l \left(u \frac{\nabla_m v}{\delta x_m} \right) \right] + \mathcal{O}(\delta x_l^4, \delta x_m^4), \quad l, m = 1, 2, 3, \quad m \neq l \quad (58)$$

with the 4th-order standard Padé operator:

$$\nabla_m = (I + \frac{1}{6}\delta_m^2)^{-1} \delta_m \mu_m$$

More precisely, the time integration (50) is now applied to

$$F_k = -\frac{\nabla_1^{(6)} f_1^E}{\delta x_1} - \frac{\nabla_2^{(6)} f_2^E}{\delta x_2} - \frac{\nabla_3^{(6)} f_3^E}{\delta x_3} + \frac{\chi_k}{2} [\delta_1(\Phi_1 \check{r}_1) + \delta_2(\Phi_2 \check{r}_2) + \delta_3(\Phi_3 \check{r}_3)] + \frac{\bar{\nabla}_1 \tilde{f}_1^V}{\delta x_1} + \frac{\bar{\nabla}_2 \tilde{f}_2^V}{\delta x_2} + \frac{\bar{\nabla}_3 \tilde{f}_3^V}{\delta x_3} \quad (59)$$

still using the numerical dissipation of the Euler scheme. In the Euler terms, the superscript ⁽⁶⁾ denotes the accuracy order of the Padé derivatives and distinguishes them from the 4th-order Padé derivatives $\frac{\bar{\nabla}_m}{\delta x_m}$ used in the viscous terms. The numerical viscous-flux in each direction are:

$$\tilde{f}_1^V = \frac{1}{Re} \left(0, \tilde{\tau}_{11}, \tilde{\tau}_{12}, \tilde{\tau}_{13}, \tilde{\psi}_1 \right)^T \quad \tilde{f}_2^V = \frac{1}{Re} \left(0, \tilde{\tau}_{21}, \tilde{\tau}_{22}, \tilde{\tau}_{23}, \tilde{\psi}_2 \right)^T \quad \tilde{f}_3^V = \frac{1}{Re} \left(0, \tilde{\tau}_{31}, \tilde{\tau}_{32}, \tilde{\tau}_{33}, \tilde{\psi}_3 \right)^T$$

They are respectively defined at $(j_1 + \frac{1}{2}, j_2, j_3)$, $(j_1, j_2 + \frac{1}{2}, j_3)$ and $(j_1, j_2, j_3 + \frac{1}{2})$ with

$$\begin{aligned} \tilde{\tau}_{11} &= (\bar{\mu}_1 \nu) \left[\frac{4}{3} \frac{\bar{\nabla}_1 u_1}{\delta x_1} - \frac{2}{3} \bar{\mu}_1 \left(\frac{\nabla_2 u_2}{\delta x_2} + \frac{\nabla_3 u_3}{\delta x_3} \right) \right], \quad \tilde{\tau}_{12} = (\bar{\mu}_1 \nu) \left(\frac{\bar{\nabla}_1 u_2}{\delta x_1} + \bar{\mu}_1 \frac{\nabla_2 u_1}{\delta x_2} \right), \\ \tilde{\tau}_{13} &= (\bar{\mu}_1 \nu) \left(\frac{\bar{\nabla}_1 u_3}{\delta x_1} + \bar{\mu}_1 \frac{\nabla_3 u_1}{\delta x_3} \right) \\ \tilde{\tau}_{21} &= (\bar{\mu}_2 \nu) \left(\frac{\bar{\nabla}_2 u_1}{\delta x_2} + \bar{\mu}_2 \frac{\nabla_1 u_2}{\delta x_1} \right), \quad \tilde{\tau}_{22} = (\bar{\mu}_2 \nu) \left[\frac{4}{3} \frac{\bar{\nabla}_2 u_2}{\delta x_2} - \frac{2}{3} \bar{\mu}_2 \left(\frac{\nabla_3 u_3}{\delta x_3} + \frac{\nabla_1 u_1}{\delta x_1} \right) \right], \\ \tilde{\tau}_{23} &= (\bar{\mu}_2 \nu) \left(\frac{\bar{\nabla}_2 u_3}{\delta x_2} + \bar{\mu}_2 \frac{\nabla_3 u_2}{\delta x_3} \right) \\ \tilde{\tau}_{31} &= (\bar{\mu}_3 \nu) \left(\frac{\bar{\nabla}_3 u_1}{\delta x_3} + \bar{\mu}_3 \frac{\nabla_1 u_3}{\delta x_1} \right), \quad \tilde{\tau}_{32} = (\bar{\mu}_3 \nu) \left(\frac{\bar{\nabla}_3 u_2}{\delta x_3} + \bar{\mu}_3 \frac{\nabla_2 u_3}{\delta x_2} \right), \\ \tilde{\tau}_{33} &= (\bar{\mu}_3 \nu) \left[\frac{4}{3} \frac{\bar{\nabla}_3 u_3}{\delta x_3} - \frac{2}{3} \bar{\mu}_3 \left(\frac{\nabla_1 u_1}{\delta x_1} + \frac{\nabla_2 u_2}{\delta x_2} \right) \right] \end{aligned}$$

and

$$\tilde{\psi}_l = (\bar{\mu}_l u_1) \tilde{\tau}_{l1} + (\bar{\mu}_l u_2) \tilde{\tau}_{l2} + (\bar{\mu}_l u_3) \tilde{\tau}_{l3} + \frac{\gamma}{Pr} (\bar{\mu}_l \nu) \frac{\bar{\nabla}_l e}{\delta x_l}, \quad l = 1, 2, 3$$

Since this scheme is spatially accurate at order 5 for the Euler terms and at order 4 for the viscous terms, it will be called *E5V4*.

The *E5V4* scheme requires only the solution of ordinary tridiagonal linear systems (TLS) with constant coefficients. Let us count these TLS on a cubic J^3 -mesh.

For the Euler terms:

– $3J^2$ TLS are necessary to compute each component of $\frac{\nabla_1 f_1^E}{\delta x_1}$, $\frac{\nabla_2 f_2^E}{\delta x_2}$ and $\frac{\nabla_3 f_3^E}{\delta x_3}$. This makes much less work than $15J^2$ TLS because our TLS algorithm takes advantage of having the same tridiagonal matrix for different right-hand sides (here 5 RHS).

For the viscous terms:

– $3J^2$ TLS are necessary to compute the standard Padé derivatives $\frac{\nabla_1 U}{\delta x_1}$, $\frac{\nabla_2 U}{\delta x_2}$ and $\frac{\nabla_3 U}{\delta x_3}$ for each of the 3 components of $U = (u_1, u_2, u_3)^T$,

– $3J^2$ TLS are necessary to compute the midpoints Padé derivatives $\frac{\bar{\nabla}_1 V}{\delta x_1}$, $\frac{\bar{\nabla}_2 V}{\delta x_2}$ and $\frac{\bar{\nabla}_3 V}{\delta x_3}$ for each of the 4 components of $V = (u_1, u_2, u_3, e)^T$,

– $3J^2$ TLS are necessary to compute the Padé averages $\bar{\mu}_1 W_1$, $\bar{\mu}_2 W_2$ and $\bar{\mu}_3 W_3$ for each of the 7 components of $W_1 = (\nu, \frac{\nabla_2 u_2}{\delta x_2} + \frac{\nabla_3 u_3}{\delta x_3}, \frac{\nabla_2 u_1}{\delta x_2}, \frac{\nabla_3 u_1}{\delta x_3}, u_1, u_2, u_3)^T$ or $W_2 = (\nu, \frac{\nabla_3 u_3}{\delta x_3} + \frac{\nabla_1 u_1}{\delta x_1}, \frac{\nabla_3 u_2}{\delta x_3}, \frac{\nabla_1 u_2}{\delta x_1}, u_1, u_2, u_3)^T$ or $W_3 = (\nu, \frac{\nabla_1 u_1}{\delta x_1} + \frac{\nabla_2 u_2}{\delta x_2}, \frac{\nabla_1 u_3}{\delta x_1}, \frac{\nabla_2 u_3}{\delta x_2}, u_1, u_2, u_3)^T$,

– $3J^2$ TLS are finally necessary to compute the midpoints Padé derivatives $\frac{\bar{\nabla}_1 \tilde{f}_1^V}{\delta x_1}$, $\frac{\bar{\nabla}_2 \tilde{f}_2^V}{\delta x_2}$ and $\frac{\bar{\nabla}_3 \tilde{f}_3^V}{\delta x_3}$ for each of the 4 non-zero components of the viscous fluxes.

Clearly, the viscous terms are more expensive to evaluate than the inviscid terms. The CPU-times per time step and mesh point for an Euler computation (*E5 – RKo6*) and a Navier–Stokes computation (*E5V4 – RKo6*) of the same 3-D problem can be found in Table 3. The CPU-time ratio is $5.17/1.91 \approx 2.7$.

5. Numerical validations

The new compact scheme has been applied to several two and three-dimensional problems using the general 3-D code.

5.1. Advection of a vortex during a long time

We first consider a two-dimensional problem governed by the Euler equations: the advection of a vortex during a long time. The vortex is homentropic and initially located at the origin ($x_1 = x_2 = 0$). The initial velocity components u_1 , u_2 and the absolute temperature T are defined in non-dimensional form as:

$$\begin{aligned} u_1 &= a_1 - \frac{\Gamma}{2\pi} x_2 \exp\left(\frac{1-r^2}{2}\right) & u_2 &= a_2 + \frac{\Gamma}{2\pi} x_1 \exp\left(\frac{1-r^2}{2}\right) \\ T &= 1 - \frac{(\gamma-1)\Gamma^2}{8\gamma\pi^2} \exp(1-r^2) \end{aligned} \quad (60)$$

with the squared radius $r^2 = x_1^2 + x_2^2$, the advection speed (a_1, a_2) and the vortex strength $\Gamma = 5$. The thermodynamic equation of state is the ideal law $p = \rho T$, with constant specific heats of ratio $\gamma = 1.4$. The uniformity of entropy gives $\rho = T^{1/\gamma-1}$. The computational domain is $[-5, 5]^2$. It is periodic in both directions.

a) Horizontal advection of the vortex

The vortex is advected in the x_1 -direction at speed $(a_1, a_2) = (0.5, 0)$. Calculations are done until time $t = 100$, after which the vortex has traveled a distance of 50. This problem was computed in [13] by the *RBC5 – ABM4* scheme on a 50×50 regular-mesh ($\delta x_1 = \delta x_2 = \delta x$) with $\Delta t = 1/50$ (maximal value for stability), which corresponds to $CFL = 0.24$, where

$$CFL = \frac{\Delta t}{\delta x} \max_j (|\vec{U}| + c_s)_j \quad (61)$$

with $|\vec{U}| = (u_1^2 + u_2^2)^{1/2}$, $c_s = (\gamma p / \rho)^{1/2}$ and $j = (j_1, j_2)$. The numerical solution of the *RBC5 – ABM4* scheme at $t = 100$ is recalled in Fig. 2. This figure compares the isobar lines computed at time $t = 100$ to the exact ones (those at $t = 0$ owing to periodicity). It also presents a pressure cut on the axis $x_2 = 0$ at $t = 0$ and $t = 100$.

This test-case has been computed with the new scheme at order 5 in space (*E5 – RKo6*) with $\chi_6 = 0.2$, using the same mesh and the same time step as for *RBC5 – ABM4*. The numerical solution is shown on Fig. 3. It appears a little more accurate than the one given by the *RBC5 – ABM4* scheme. Furthermore, the new scheme is less expensive by time iteration (in CPU time and also in memory requirements) and allows much larger time-steps as shown in Table 2. Note that the solution of *E5 – RKo6* at $CFL = 1$ cannot be distinguished from that at $CFL = 0.24$ at the scale of Fig. 3.

b) Diagonal advection of the vortex

The vortex is now advected in the diagonal direction at speed $(a_1, a_2) = (0.5, 0.5)$ in the same periodic domain. Calculations are done again until time $t = 100$, after which the vortex has traveled a distance of $50\sqrt{2}$. The solution of *E5 – RKo6* on

Table 2

Computational cost on a single core of Westmere Intel-X5680 for the vortex advection during a long time on a 50×50 mesh (using a 3-D code for the new scheme $E5 - RKo6$).

	$RBC_5 - ABM_4$	$E5 - RKo6$	$E5 - RKo6$
Mesh	50×50	$50 \times 50 \times 3$	$50 \times 50 \times 3$
CFL	0.24	0.24	1
Time iterations for $t = 100$	5000	5000	1210
CPU time (s.) for $t = 100$	91.4	49.4	12, 04

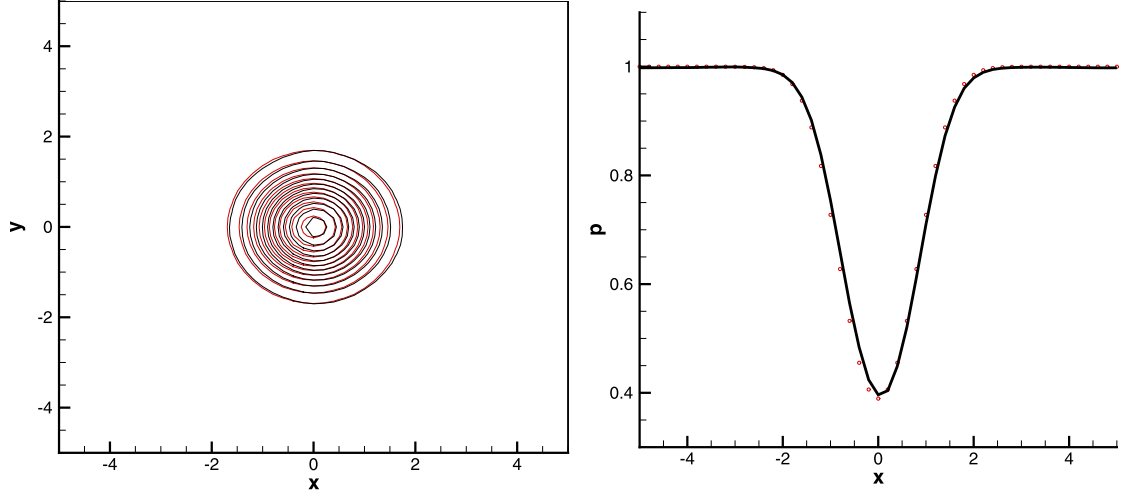


Fig. 2. Horizontal advection of a vortex by $RBC_5 - ABM_4$ scheme [13] on a 50×50 mesh at $CFL = 0.24$ with periodicity conditions. Left: isobars (from $p = 0.4$ to $p = 0.95$, $\Delta p = 0.05$) at $t = 0, 100$. Right: pressure cut on the axis $x_2 = 0$ at $t = 0$ (symbols) and at $t = 100$ (line).

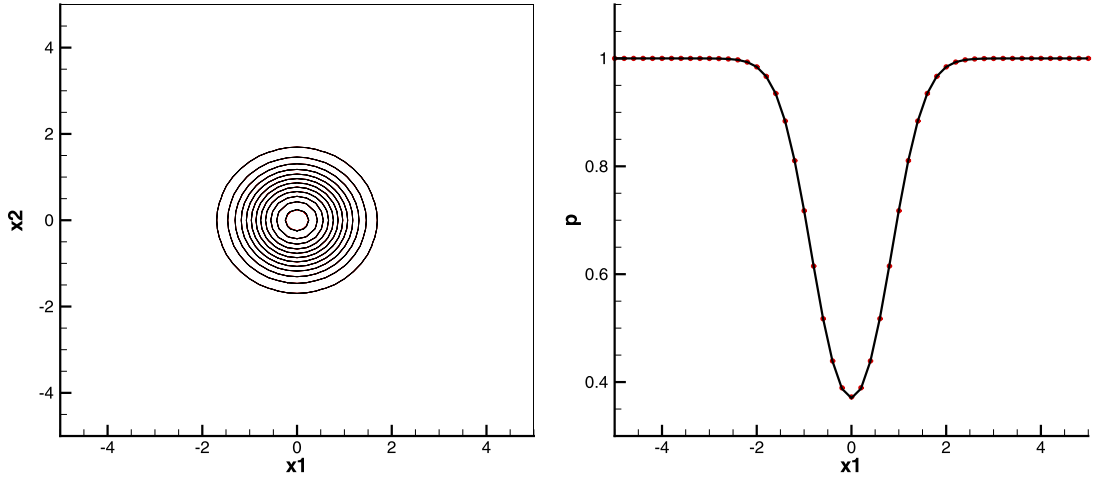


Fig. 3. Horizontal advection of a vortex by $E5 - RKo6$ scheme on a 50×50 mesh at $CFL = 0.24$ with periodicity conditions. Left: isobars (from $p = 0.4$ to $p = 0.95$, $\Delta p = 0.05$) at $t = 0, 100$. Right: pressure cut on the axis $x_2 = 0$ at $t = 0$ (symbols) and at $t = 100$ (line).

a 50×50 mesh at $CFL = 1$ is shown on Fig. 4. This solution is as accurate as that of the horizontal advection. This can be checked more precisely on Fig. 5 showing the L_2 errors on pressure at $t = 100$ for the horizontal and diagonal advection of the vortex. These errors are plotted versus the mesh size ($h = \delta x$) for $CFL = 1/4, 1/2$ and 1. Their slopes in log-log scale give the accuracy order at $t = 100$. For the first two CFL -numbers, this order is close to 5, i.e. to the space accuracy-order. For $CFL = 1$, it becomes close to 4 because of the time-integration error.

5.2. 3D-diagonal advection of a spherical Gaussian

Consider the diagonal advection of a spherical Gaussian shape in a cubic domain:

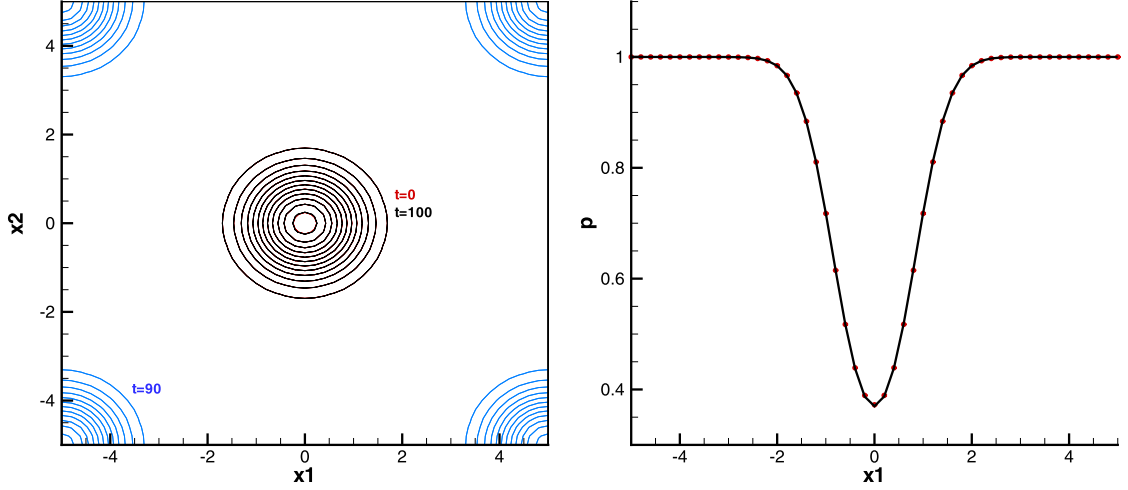


Fig. 4. Diagonal advection of a vortex by $E5 - RKo6$ scheme on a 50×50 mesh at $CFL = 1$ with periodicity conditions. Left: isobars (from $p = 0.4$ to $p = 0.95$, $\Delta p = 0.05$) at $t = 0, 90, 100$. Right: pressure cut on the axis $x_2 = 0$ at $t = 0$ (symbols) and at $t = 100$ (line).

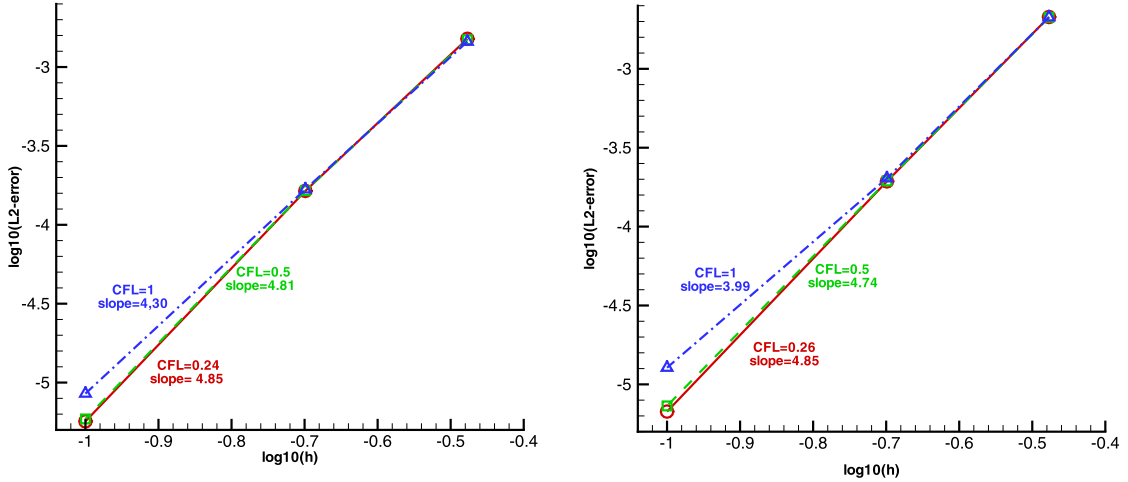


Fig. 5. L_2 errors on pressure at $t = 100$ for the advection of a vortex by $E5 - RKo6$ scheme on a series of meshes (30×30 , 50×50 , 100×100). Left: horizontal advection. Right: diagonal advection.

$$\begin{cases} \frac{\partial w}{\partial t} + A_1 \frac{\partial w}{\partial x_1} + A_2 \frac{\partial w}{\partial x_2} + A_3 \frac{\partial w}{\partial x_3} = 0 \\ w(x_1, x_2, x_3, 0) = \exp(-75 r^2), \quad (x_1, x_2, x_3) \in [-1, 1]^3 \end{cases} \quad (62)$$

with $r^2 = (x_1^2 + x_2^2 + x_3^2)$, $A_1 = A_2 = A_3 = 1$ and periodicity in the three directions.

The spherical Gaussian is initially centered at the origin and moves along the diagonal line $x_1 = x_2 = x_3$ at speed $\sqrt{3}$. This problem is solved on a uniform Cartesian mesh ($\delta x_1 = \delta x_2 = \delta x_3 = \delta x$) until time $t = 2$, after which the Gaussian has traveled a distance of $2\sqrt{3}$ and has returned to its initial location. For this diagonal-advection problem, the stability condition (55) becomes:

$$3 \frac{\Delta t}{\delta x} \leq \eta$$

or else $CFL \leq \eta/\sqrt{3}$ using the usual definition:

$$CFL = \frac{\Delta t}{\delta x} \sqrt{A_1^2 + A_2^2 + A_3^2} = \frac{\Delta t}{\delta x} \sqrt{3}$$

We use the $E5 - RKo6$ scheme with $\chi_6 = 0.2$, for which $\eta = 2$, so that the stability condition is $CFL \leq 2/\sqrt{3}$. Fig. 6 displays the numerical solution for $CFL_{max} = 2/\sqrt{3}$ on a 100^3 -mesh. The L_2 errors on w versus the mesh size are plotted at time $t = 2$ on Fig. 7 for CFL_{max} , $CFL_{max}/2$ and $CFL_{max}/4$. The accuracy orders obtained are similar to those found for the 2D vortex-advection problems and leads to the same conclusions.

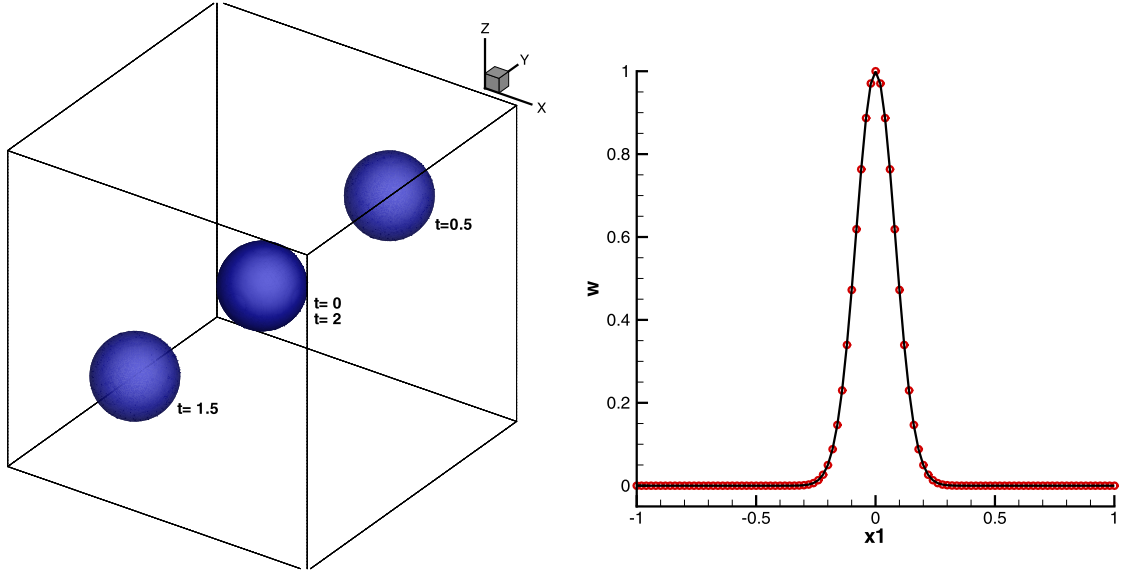


Fig. 6. 3D-diagonal advection of a spherical Gaussian by $E5 - RKo6$ scheme on a 100^3 -mesh at $CFL = 2/\sqrt{3}$. Left: iso-surface $w = 0.01$ at $t = 0, 0.5, 1.5, 2$. Right: x_1 -profile in plane $x_2 = x_3 = 0$ at $t = 0$ (mesh points) and $t = 2$ (solid line).

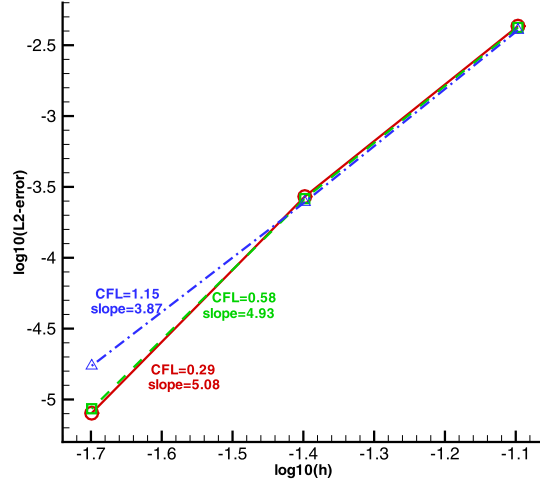


Fig. 7. 3D-diagonal advection of a spherical Gaussian by $E5 - RKo6$ scheme: L^2 -errors at $t = 2$ on a series of meshes ($25^3, 50^3, 100^3$) for $CFL = 2/\sqrt{3}, 1/\sqrt{3}$ and $0.5/\sqrt{3}$.

5.3. DNS of the Taylor–Green Vortex

We now consider the dimensionless Navier–Stokes equations (56) and apply the new compact scheme to the direct numerical simulation of the Taylor Green Vortex in the three-dimensional periodic domain $[0, 2\pi]^3$ starting with the initial conditions:

$$\begin{cases} u_1(x_1, x_2, x_3, 0) = \sin(x_1) \cos(x_2) \cos(x_3) \\ u_2(x_1, x_2, x_3, 0) = -\cos(x_1) \sin(x_2) \cos(x_3) \\ u_3(x_1, x_2, x_3, 0) = 0 \\ p(x_1, x_2, x_3, 0) = p_0 + \frac{1}{16}[\cos(2x_1) + \cos(2x_2)][\cos(2x_3) + 2] \\ \rho(x_1, x_2, x_3, 0) = p(x_1, x_2, x_3, 0)/p_0 \end{cases} \quad (63)$$

where $p_0 = 1/(\gamma M_0^2)$ with $M_0 = 0.1$ and $\gamma = 1.4$. The Reynolds and Prandtl numbers are $Re = 1600$, $Pr = 0.71$ and the viscosity coefficients do not depend on temperature. This test case was proposed by the International Workshop on High-Order CFD Methods [24].

Simulations have been run for 20 convective times ($t = 20$) on three Cartesian meshes ($64^3, 128^3, 256^3$). The time step has been calculated every 10 time-iterations from the condition $CFL = 1$, with a CFL number defined as if the flow were

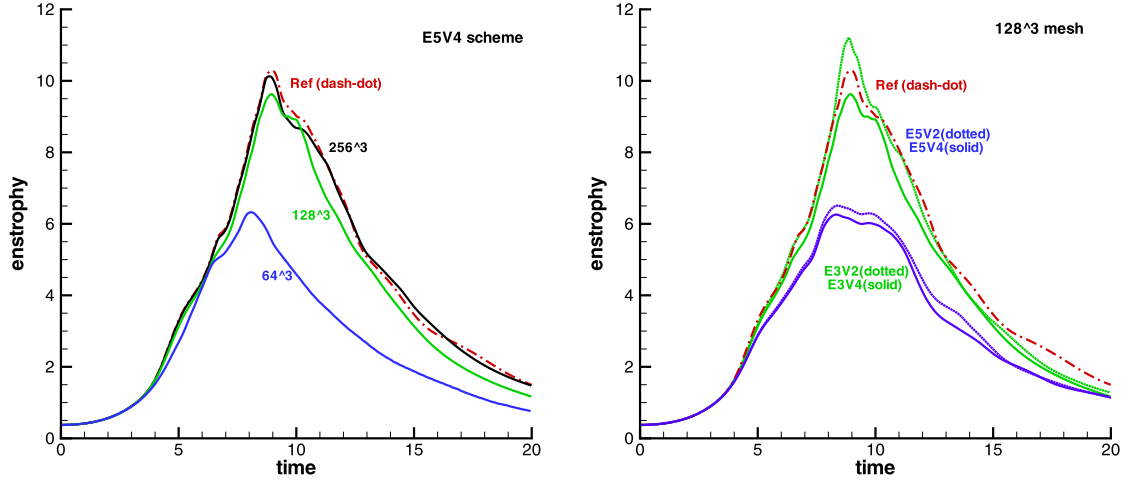


Fig. 8. Enstrophy evolution of the Taylor-Green Vortex at $CFL = 1$. Left: $E5V4 - RKo6$ solution in a series of meshes (64^3 , 128^3 , 256^3) and reference solution. Right: effect of the accuracy-order in the approximation of Euler or viscous terms ($E3V2$, $E3V4$, $E5V2$, $E5V4$) in the 128^3 -mesh.

inviscid (3D version of (61)). For this type of problem, the physical quantity which is the most sensitive for assessing the simulation accuracy is the enstrophy:

$$\epsilon = \frac{1}{(2\pi)^3} \int_{[0,2\pi]^3} \frac{1}{2} \rho |\vec{\omega}|^2 dx_1 dx_2 dx_3$$

where $\vec{\omega} = \text{rot} \vec{U}$ is the vorticity vector. For computing $\vec{\omega}$, the spatial derivatives of the velocity components have been approximated by the same standard Padé formulas (18) as those used in the Euler part of the scheme (for instance, 6th-order Padé formulas are used to compute the vorticity of the $E5V4$ solution).

On the left side of Fig. 8, the temporal evolution of the enstrophy obtained by the $E5V4 - RKo6$ scheme is plotted for the three meshes and compared to the reference solution provided by the Workshop [24]. The solution in the fine mesh (256^3), computed with $\chi_6 = 0.1$, is very close to the reference solution (see also the right side of Fig. 9). The solution in the 128^3 mesh is quite good. The solution in the 64^3 mesh is far from the reference solution, but better than in many contributions to the Workshop.

In order to evaluate the effect of the accuracy order on the approximation of Euler terms and viscous terms, we have run four simulations on the medium mesh with the spatial orders: $E3V2$, $E3V4$, $E5V2$ and $E5V4$. The results (see the right side of Fig. 8) demonstrate that using a 3rd-order treatment of the inviscid terms is quite insufficient: the quality of the solution is not better than that obtained in the 64^3 mesh with the 5th-order approximation. Moreover 3rd and 5th-order treatments have comparable costs since they both require the solution of tridiagonal linear systems. The effect of the accuracy order on the viscous terms is real but not as important as on the inviscid terms (compare $E5V2$ and $E5V4$ in the medium mesh on Fig. 8 and also in the fine mesh on Fig. 9). Besides, contrary to the effect on inviscid terms, raising the accuracy order of the viscous terms decreases the enstrophy peak, i.e. increases the dissipation. This can be easily understood by expanding the approximations of orders 2 and 4 of a second derivative. For a constant viscosity ν ($\nu = 1$ in dimensionless form), in a direction x_l we obtain:

$$\frac{\delta_l}{\delta x_l} \left(\nu \frac{\delta_l v}{\delta x_l} \right) = \frac{\partial^2 v}{\partial x_l^2} + \frac{\delta x_l^2}{12} \frac{\partial^4 v}{\partial x_l^4} + \mathcal{O}(\delta x_l^4) \quad (64)$$

$$\frac{\nabla_l}{\delta x_l} \left(\nu \frac{\nabla_l v}{\delta x_l} \right) = \left[\left(I + \frac{1}{24} \delta_l^2 \right) \left(I + \frac{1}{24} \delta_l^2 \right) \right]^{-1} \frac{\delta_l^2 v}{\delta x_l^2} = \frac{\partial^2 v}{\partial x_l^2} - \frac{17}{2880} \delta x_l^4 \frac{\partial^6 v}{\partial x_l^6} + \mathcal{O}(\delta x_l^6) \quad (65)$$

The second derivative of v is of course dissipative, but the errors, of order 2 in (64) and of order 4 in (65), are both anti-dissipative. Thus, replacing the anti-dissipative error in (64) by a smaller one in (65) increases the overall dissipation. It can be checked that this is also true for the approximation of a crossed second-derivative. Anyway, the solution given by $E5V4$ in the fine mesh is slightly better than that of $E5V2$, especially in the enstrophy decay-phase dominated by viscous diffusion (see Fig. 9).

The temporal evolution of the total kinetic energy:

$$E_k = \frac{1}{(2\pi)^3} \int_{[0,2\pi]^3} \frac{1}{2} \rho |\vec{U}|^2 dx_1 dx_2 dx_3$$

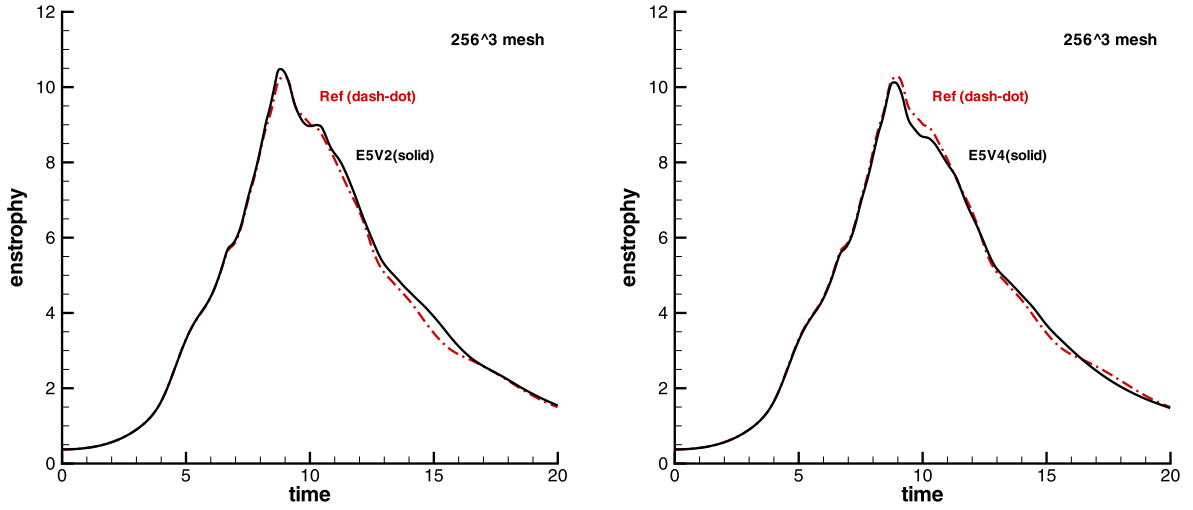


Fig. 9. Enstrophy evolution of the Taylor–Green Vortex in the 256^3 -mesh at $CFL = 1$: effect of the accuracy order in the approximation of viscous terms. Left: $E5V2$ solution and reference solution. Right: $E5V4$ solution and reference solution.

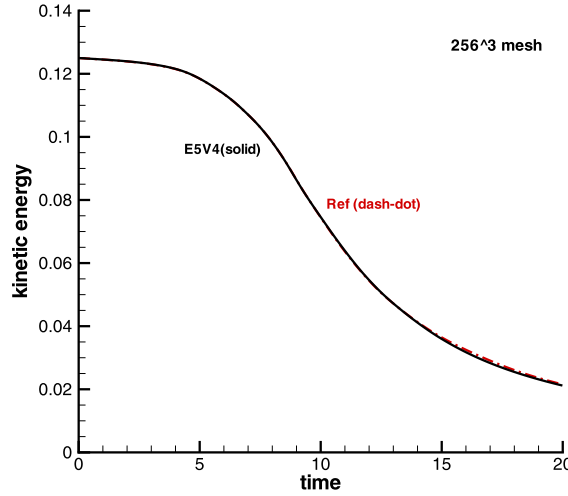


Fig. 10. Kinetic-energy evolution of the Taylor–Green Vortex for the $E5V4$ solution in the 256^3 -mesh at $CFL = 1$ compared to the reference solution.

Table 3

Computational costs at $CFL = 1$ on Westmere Intel-X5680 for the Taylor–Green Vortex.

Scheme	pts	t	n (time it.)	Cores	CPU-time	CPU-time/ n	CPU-time/ n /pts
$E5 - RKo6$ (Euler)	64^3	6	678	1	340.2 s	0.502 s	1.91 μ s
$E5V4 - RKo6$	64^3	20	2238	1	3035.7 s	1.356 s	5.17 μ s
$E5V4 - RKo6$	128^3	20	4462	6	10583 s (2 h 56 min)	2.37 s	1.13 μ s
$E5V4 - RKo6$	256^3	20	8916	12	100852 s (28 h)	11.31 s	0.67 μ s

is shown on Fig. 10 for the $E5V4 - RKo6$ scheme on the fine mesh. It is very close to reference solution. This is also true on the medium mesh. Computing times of the present method are listed in Table 3. A rough parallelization of the code using OpenMP has been done for the computations on the medium and fine meshes.

5.4. Shock–vortex interaction

To assess the shock-capturing capabilities of the new scheme, we consider a moving homentropic vortex interacting with a steady shock. This problem was proposed by Jiang and Shu [25]. The space domain is $[0, 1]^2$. The vortex is initially located at $(x_c = 0.25, y_c = 0.5)$ and advected in the x -direction. The shock is positioned at $x = 0.5$ and normal to the x -axis.

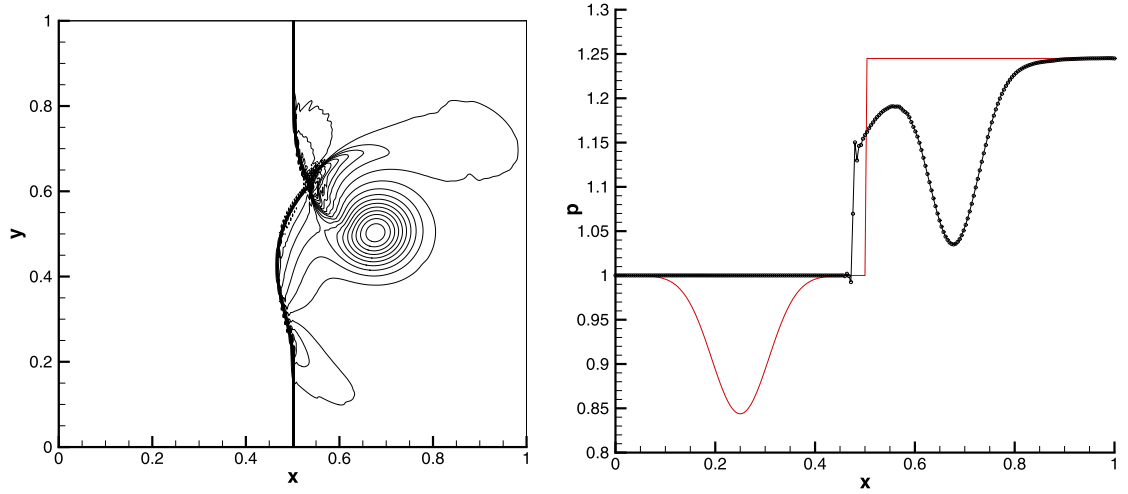


Fig. 11. Shock-vortex interaction computed by $RBC_5 - ABM_4$ scheme [13] on a 248×248 mesh at $CFL = 0.17$. Left: isobar lines (from $p = 0.79$ to $p = 1.61$, $\Delta p = 0.02$) at $t = 0.35$. Right: pressure cut on $y = 0.5$ at $t = 0$, and at $t = 0.35$ (line with mesh points).

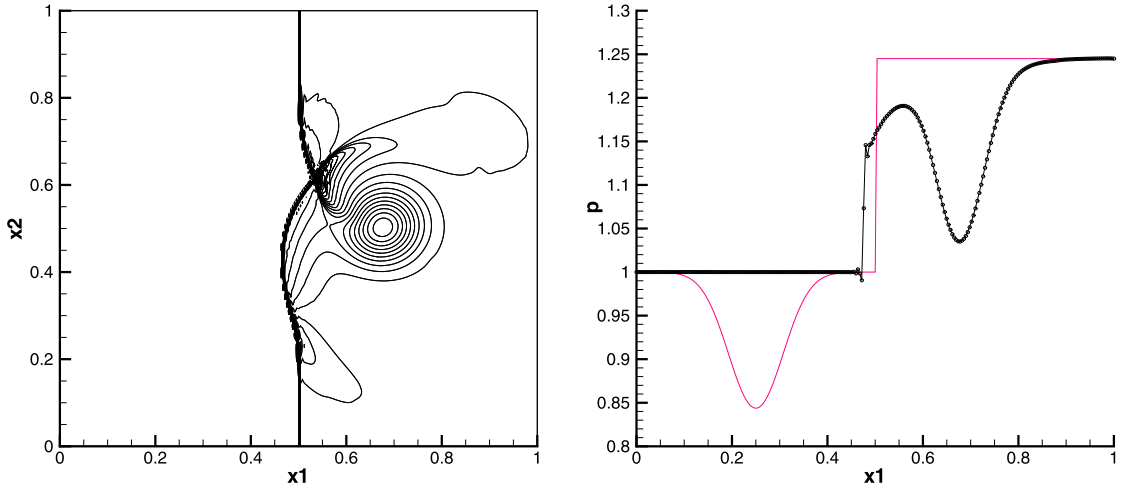


Fig. 12. Shock-vortex interaction computed by $E5 - RKo6$ scheme on a $248 \times 248 \times 3$ mesh at $CFL = 0.5$. Left: isobar lines (from $p = 0.79$ to $p = 1.61$, $\Delta p = 0.02$) at $t = 0.35$. Right: pressure cut on $y = 0.5$ at $t = 0$ and at $t = 0.35$ (line with mesh points).

Its upstream Mach number is $M_0 = 1.1$. The initial velocity components u , v and the absolute temperature T are defined upstream the shock in non-dimensional form as:

$$u = u_0 + \epsilon Y \exp[\alpha(1 - R^2)] \quad v = -\epsilon X \exp[\alpha(1 - R^2)]$$

$$T = 1 - \frac{(\gamma - 1)\epsilon^2}{4\alpha\gamma} \exp[2\alpha(1 - R^2)]$$

with

$$X = \frac{x - x_c}{r_c}, \quad Y = \frac{y - y_c}{r_c}, \quad R^2 = X^2 + Y^2$$

where $\epsilon = 0.3$ (vortex strength), $\alpha = 0.204$ (decay-rate control), $r_c = 0.05$ (radius of maximum vortex-strength) and $u_0 = M_0 \sqrt{\gamma}$. The gas law is $p = \rho T$, with $\gamma = 1.4$. Downstream the initial shock, the flow is uniform at the subsonic conditions behind the steady normal shock.

On the boundary the following conditions are applied: supersonic inflow on left side, subsonic outflow on right side and solid wall on lateral sides. Calculations are made on a 248×248 regular-mesh until the time $t = 0.35$. Results obtained in [13] by the $RBC_5 - ABM_4$ scheme are recalled in Fig. 11. They were obtained with a constant time-step $\Delta t = 0.35/1400$ (the time-integration was multi-level) corresponding to a CFL number of about 0.17. The present solution given by the $E5 - RKo6$ scheme with $\chi_6 = 2$ is shown on Fig. 12. It has been computed by the 3-D code using a $248 \times 248 \times 3$ mesh with periodicity

in the third direction, at $CFL = 0.5$ which requires 486 time-steps (average $\Delta t = 0.35/486$). The two solutions are close and produce a sharp shock-wave spread over two mesh cells. The present scheme is a little less oscillatory and is 7 times quicker than the $RBC_5 - ABM_4$ scheme. Note that these high-order simulations have been run without limiters, entropy correction, filtering or other additives.

6. Conclusion

A new compact method has been proposed for the unsteady compressible Euler and Navier–Stokes equations. This method preserves the best features of the RBC approximation, notably its type of numerical dissipation. A high accuracy at moderate cost has been obtained by using compact formulas of order 6 for the Euler terms and of order 4 for the viscous terms (even if the viscosity coefficients vary). These formulas only need the solution of tridiagonal linear systems with constant coefficients. The use of the Bogey–Bailly time integration requires a minimum of computer memory and allows a CFL number equal to one or slightly greater, for 3-D problems. The numerical dissipation requires no additional linear algebra and involves matrix-coefficients Φ_l simpler than those used in the RBC schemes. Thus, the computation of the numerical dissipation is pretty cheap and moreover done only once per time-step, at the last Runge–Kutta stage.

Additional studies are still needed to improve the shock capturing properties of the method. Corrections could be inspired by the discrete shock analysis of the RBC schemes presented in [26,27]. Calculations of more complex flow problems are also planned.

References

- [1] S.K. Lele, Compact finite difference schemes with spectral-like resolution, *J. Comput. Phys.* 103 (1992) 16–42.
- [2] B. Cockburn, C.-W. Shu, Nonlinear stable compact schemes for shock calculations, *SIAM J. Numer. Anal.* (1994) 607–627.
- [3] H.C. Yee, Explicit and implicit multidimensional compact high-resolution shock-capturing methods: formulation, *J. Comput. Phys.* 131 (1997) 216–232.
- [4] M.R. Visbal, D.V. Gaitonde, High-order accurate methods for complex unsteady subsonic flows, *AIAA J.* 37 (1999) 1231–1239.
- [5] A.I. Tolstykh, High Accuracy Non-Centered Compact Difference Schemes for Fluid Dynamics Applications, World Scientific, Singapore, 1994.
- [6] D. Fu, Y. Ma, A high order accurate difference scheme for complex flow fields, *J. Comput. Phys.* 134 (1997) 1–15.
- [7] A. Lerat, C. Corre, A residual-based compact scheme for the compressible Navier–Stokes equations, *J. Comput. Phys.* 170 (2001) 642–675.
- [8] A. Lerat, C. Corre, Residual-based compact schemes for multidimensional hyperbolic systems of conservation laws, *Comput. Fluids* 31 (2002) 639–661.
- [9] A. Lerat, C. Corre, Higher order residual-based compact schemes on structured grids, in: 34th Comput. Fluid Dyn. Course, von Karman Institute for Fluid Dynamics, 2006, pp. 1–111, VKI LS 2006-1.
- [10] C. Corre, F. Falissard, A. Lerat, High-order residual-based compact schemes for compressible inviscid flows, *Comput. Fluids* 36 (2007) 1567–1582.
- [11] A. Lerat, K. Grimich, P. Cinnella, On the design of high order residual-based dissipation for unsteady compressible flows, *J. Comput. Phys.* 235 (2013) 32–51.
- [12] K. Grimich, P. Cinnella, A. Lerat, Spectral properties of high-order residual-based compact schemes for unsteady compressible flows, *J. Comput. Phys.* 252 (2013) 142–162.
- [13] A. Lerat, A high-order time formulation of the RBC schemes for unsteady compressible Euler equations, *J. Comput. Phys.* 303 (2015) 251–268.
- [14] R. Abgrall, Residual distribution schemes: current status and future trends, *Comput. Fluids* 35 (2006) 641–669.
- [15] H. Deconinck, M. Ricchiuto, Residual Distribution Schemes: Foundation and Analysis, Encyclopedia of Computational Mechanics, John Wiley and Sons, 2007.
- [16] R. Abgrall, A. Larat, M. Ricchiuto, C. Tavé, A simple construction of very high order non-oscillatory compact schemes on unstructured meshes, *Comput. Fluids* 38 (2009) 1314–1323.
- [17] M. Ricchiuto, R. Abgrall, Explicit Runge–Kutta residual distribution schemes for time dependent problems: second order case, *J. Comput. Phys.* 229 (2010) 5653–5691.
- [18] R. Abgrall, D.D. Santis, Linear and non-linear high order accurate residual distribution schemes for the discretization of the steady compressible Navier–Stokes equations, *J. Comput. Phys.* 283 (2015) 329–359.
- [19] C. Corre, G. Hanss, A. Lerat, A residual-based compact scheme for the unsteady compressible Navier–Stokes equations, *Comput. Fluids* 34 (2005) 561–580.
- [20] K. Grimich, B. Michel, P. Cinnella, A. Lerat, An accurate finite-volume formulation of a residual-based compact scheme for unsteady compressible flows, *Comput. Fluids* 92 (2014) 93–112.
- [21] C. Bogey, C. Bailly, A family of low dispersive and low dissipative explicit schemes for flow and noise computations, *J. Comput. Phys.* 194 (2004) 194–214.
- [22] P.L. Roe, Approximate Riemann solvers, parameter vectors and difference schemes, *J. Comput. Phys.* 43 (1981) 357–372.
- [23] Y. Huang, A. Lerat, Second-order upwinding through a characteristic time-step matrix for compressible flow calculations, *J. Comput. Phys.* 142 (1998) 445–472.
- [24] Third International Workshop on High-Order CFD Methods, Orlando, USA <http://www.grc.nasa.gov/hiocfd>, 2015.
- [25] G. Jiang, C. Shu, Efficient implementation of weighted ENO schemes, *J. Comput. Phys.* 126 (1996) 202–228.
- [26] A. Lerat, Steady discrete shocks of 5th and 7th-order RBC schemes and shock profiles of their equivalent differential equations, *J. Comput. Phys.* 272 (2014) 629–643.
- [27] A. Lerat, Shock capturing analysis of high-order schemes through analytical solution of the discrete schemes and their equivalent differential equations, in: International Conference on Computational Fluid Dynamics 8, Chengdu, China, 2014, ICCFD8-2014-198.